C8.7 Optimal Control

Sheet 1 — HT25

Section A

1. Consider a finite directed graph, that is a set of nodes $V = \{1, 2, ..., N\}$ and edges $E \subset V \times V$. We assume that all self-connections are possible, that is, $(x, x) \in E$ for all $x \in V$. Suppose the graph is connected, that is, for any $x, x' \in V$ there exists $m \in \mathbb{N}$ and a sequence $x = x_0, x_1, x_2, ..., x_m = x'$ with $(x_i, x_{i+1}) \in E$ for all i. We call such a sequence a path from x to x'.

   (a) Explain why there exists a path from x to x' with exactly N steps.

   (b) Define the cost of following a path of length $T \geq N$ to be

   $$J(U) = \Big(\sum_{t=1}^{T-1} f(X_t, U_t)\Big) + \Phi(X_T)$$

   where $U_t \in V$ determines the next step in our path (so $X^U_{t+1} = U_t$) and, for a fixed state $x^* \in V$,

   $$f(x,u) = \begin{cases} 0 & \text{if } x = x^*, \\ 1 & \text{if } (x,u) \in E, x \neq x^*, , \\ \infty & \text{if } (x,u) \notin E. \end{cases} \qquad \Phi(x) = \begin{cases} 0 & \text{if } x = x^* \\ N+1 & \text{if } x \neq x^* \end{cases}$$

   Show that $\inf_U J(t, U)$ is the length of the shortest path to x^*, and write down the dynamic programming formulation to find the shortest path. (This is a variant of Dijkstra's algorithm.)

   (c) Now suppose that at each time t, we randomize whether a connection x, x' is available. That is, for each t and $(x, x') \in E$, there is a Bernoulli random variable $Y_{t,x,x'}$, with probability p such that the path from x to x' is available if $Y_{t,x,x'} = 1$. We assume all edges are independent, and whether an edge is available is known only at time t. Express this by varying the cost function f, and write down a dynamic programming formulation to find the shortest expected path length.

Solution:

(a) If there exists a path from x to x', then there exists a path without repeated nodes. As there are N nodes in total, we know that a path without repeats will have length at most N. Now allowing repeats in the final node, we see that there must be a path with exactly N steps.

Sam Cohen: cohens@maths.ox.ac.uk

(b) As there exists a path with at most N steps, and we can repeat the final node with zero cost, we see that $J(0, U) < N$. In particular, for U minimizing $J(0, x_0, U)$, we know that $\Phi(X_T^U) < N + 1$, so $X_T^U = x^*$. In this case, the costs are simply the number of non-repeating steps in the path. The dynamic programming formula then states that the length of the shortest path is $V_0(x_0) = \min_U J(0, U)$, and we have

$$V_t(x) = \min_{u:(x,u)\in E} \Big\{ f(t, x, u) + V_{t+1}(u) \Big\}$$

with terminal value $V_T(x) = \Phi$. We also observe that $V_t(x) < N$ if and only if there is a path from x to x^* in at most $T - t$ steps.

(c) With randomization, our problem becomes a little more complicated. If we look for paths of length $N - t$ or less (and we place a penalty of N on any path which does not reach x^* in at most N steps), the dynamic programming equation becomes

$$V_t(x, \{Y_{t,x,x'}\}_{x'\in V}) = \min_{u:Y_{t,x,u}=1} \Big\{ f(t, x, u) + \mathbb{E}[V_{t+1}(u, \{Y_{t+1,u,x'}\}_{x'\in V})] \Big\}$$

(as we know the Y are independent, so we don't need conditional expectations).

This gives us our dynamic programming formalism, but with a penalty $N+1$ if we don't reach state x^* in at most $T - t$ steps. One way to extend this is to notice that if $(x, x^*) \in E$, then the number of times we have to wait until reaching x^* is (at most) a geometric random variable, and so the expected value is at most $1/p$. By extension, the expected time to get to x^* from any state x is at most N/p. If we change $\Phi(x) = N/p$ for $x \neq x^*$, then we observe that $0 \leq v_t(x, y) \leq v_{t+1}(x, y')$ for any y, y'. Consequently, we can take a limit $t \to -\infty$ (or $T \to \infty$), and v will converge to the expected length.

2. Consider an infinite horizon discounted Markov decision problem, with finitely many actions. Show that if there are $|\mathcal{U}| = m$ actions, and $|\mathcal{X}| = n$ states, then the value function can be obtained by solving m^n systems of n linear equations and taking the componentwise minimum of these solutions.

Solution: As we are in an infinite-horizon discounted MDP, we know that there exists an optimal feedback control $u : \mathcal{X} \to \mathcal{U}$. As there are m^n maps of this type, we can enumerate them fully. For each choice of feedback control, we can compute (with an abuse of notation) the cost-to-go function

$$J(x, u) = g(x, u(x)) + e^{-\rho} \sum_{x'} p(x'; x, u(x)) J(x', u).$$

With u fixed, this is a system of n linear equations (one for each value of x). If we solve this for every choice of u, we can then directly minimize to find $v(x) = \min_u J(x, u)$, where the minimization is over maps $u : \mathcal{X} \to \mathcal{U}$.

3. Consider a Markov decision problem with costs $g(t, x, u)$ and transition probabilities $p(x'; t, x, u)$.
   (a) If p is affine with respect to u, and g is strictly convex with respect to u, show that there will be a unique optimal strategy.
   (b) If p and g are both affine with respect to u, and u takes values in an interval $[u_{\min}, u_{\max}]$, show that there is an optimal strategy which only ever takes the values $u_{\min}$ and $u_{\max}$.

   Solution:

   (a) From the Bellman equation, we know that a control U^* is optimal if and only if

   $$U_t^* \in \arg\min_{u\in\mathcal{U}} \Big\{ g(t, x, u) + \sum_{x'} p(x'; t, x, u) v(t + 1, x') \Big\}.$$

   We observe that the term inside the brackets is the sum of a strictly convex function and a collection of affine functions (with respect to u), and hence is strictly convex. However, strictly convex functions have unique minimizers, so there is a unique choice of u which satisfies this property. Therefore, the optimal control is unique.

   (b) As above, we now see that U^* is optimal if and only if it is the minimizer of an affine function on an interval. Affine functions achieve their minima at the boundaries of their domains, so U^* will only take values on the boundary of $[u_{\min}, u_{\max}]$.

4. Write down a Markov decision problem with two states, for which there *exists* an optimal non-Markovian solution.

   Solution: The key here is to find any problem where, at some time after the first, there is more than one optimal action that can be taken. You can then choose the strategy at time $t = 3$ (for example) to depend on the state of the system at time $t = 1$, so the control is optimal (as it takes values within the optimal set) but not Markov. A trivial example is any system where the control has no impact, but can take more than one value.

Sam Cohen: cohens@maths.ox.ac.uk

5. Consider a deterministic control problem, where at each time t, our agent's preferences are described by a cost-to-go function

$$J(t,x,U) = \sum_{s=t}^{T} \frac{1}{1+(s-t)} g(s,X_s,U_s).$$

(This is sometimes known as *hyperbolic discounting*). We assume $f(t,x,u) = x+u$ and $g(t,x,u) = -u + x^2$, where $x_0 \in [0,1]$ and $u \in \mathcal{U} = [0,2]$, with a horizon $T = 2$.

Show that the dynamic programming principle fails here – there are controls U which optimize $J(t, X_t, U)$ but do not optimize $J(t+1, X_{t+1}, U)$. (This is known as a *time-inconsistent problem*, as we have no DPP.)

Solution: We write the control $U = U_0, U_1, U_2$, so that

$$\begin{aligned}
J(2, X_2, U_2) &= -U_2 + X_2^2 \\
J(1, X_1, (U_1, U_2)) &= \big(-U_1 + X_1^2\big) + \frac{1}{2}\big(-U_2 + (X_1+U_1)^2\big) \\
J(0, X_0, (U_0, U_1, U_2)) &= \big(-U_0 + X_0^2\big) + \frac{1}{2}\big(-U_1 + (X_0+U_0)^2\big) \\
&\quad + \frac{1}{3}\big(-U_2 + (X_0+U_0+U_1)^2\big)
\end{aligned}$$

Observe that all of these functions are convex in U_0, U_1, U_2, as appropriate, so we expect unique minimizers. We then solve for the minimizing controls in each case:

- Optimizing $J(2,\cdot)$ gives $U_2 = 2$.
- Optimizing $J(1,\cdot)$ gives $U_2 = 2$, and $U_1 = \begin{cases} 0 & X_1 > 1 \\ 1 - X_1 & X_1 \in [-1,1] \\ 2 & X_1 < -1 \end{cases}$.
- Differentiating $J(0,\cdot)$ gives (assuming an interior solution) the system of equations

$$\begin{aligned}
U_2 &= 2 \\
0 &= \frac{1}{2} + \frac{1}{3}\big(-2(X_0+U_0+U_1)\big) \\
0 &= 1 + \frac{1}{2}(-2(X_0+U_0)) + \frac{1}{3}\big(-2(X_0+U_0+U_1)\big)
\end{aligned}$$

  from which we see $U_0 = \frac{1}{2} - X_0$ and $U_1 = \frac{5}{4}$. In particular the unique optimal strategy at time 0 has $U_1 = 5/4$, which is not optimal at time 1, except in the special case where $X_1 = -1/4$. However, this state cannot occur if we start with strategy $U_0 = \frac{1}{2} - X_0$, as this will invariably lead to $X_1 = 1/2$, no matter what the initial position is.

Sam Cohen: cohens@maths.ox.ac.uk

Section B

6. Consider a Markov decision problem, where we seek to minimize

$$J(t,x,U) = \mathbb{E}\Big[\sum_{s=t}^{T-1} g(s, X_s^{t,x,U}, U_s) + \Phi(X_T^{t,x,U})\Big|\mathcal{F}_t\Big].$$

Show that any optimal control for this problem is also optimal for the problem of minimizing

$$\tilde{J}(t,x,U) = \mathbb{E}\Big[\sum_{s=t}^{T-1}\Big(g(s, X_s^{t,x,U}, U_s) + h(s)\Big) + \Big(\Phi(X_T^{t,x,U}) + h(T)\Big)\Big|\mathcal{F}_t\Big]$$

where $h : \mathbb{T} \to \mathbb{R}$ is a deterministic function. Hence show that for any optimal control problem, there exists an equivalent problem (i.e. one with the same optimal strategies) such that $\max_x \tilde{v}(t,x) = 0$, for all $x \in \mathcal{X}$.

Solution: We can write down the DPP formulation for $\tilde{J}(t,x,U)$, which gives us

$$\tilde{v}(t,x) = \min_U \mathbb{E}\Big[\tilde{g}(t,x,U)+\tilde{v}(t+1, X_{t+1}^{t,x,U})\Big|\mathcal{F}_t\Big] = \min_u \Big\{\tilde{g}(t,x,u)+\sum_{x'} p(x';x,u)\tilde{v}(t+1,x')\Big\}$$

where $\tilde{g} = g - h$, and $\tilde{v}_T(\cdot) = \Phi(\cdot) - h(T)$. We now proceed by induction: for $t = T$, we take $h(T) = -\min_x \Phi(x)$, so that $\max_x \tilde{v}_T(x) = 0$. Inductively, we can then write

$$\tilde{v}(t,x) = \min_u \Big\{g(t,x,u) + \sum_{x'} p(x';x,u)\tilde{v}(t+1,x')\Big\} + h(t)$$

so taking

$$h(t) = -\max_x \min_u \Big\{g(t,x,u) + \sum_{x'} p(x';x,u)\tilde{v}(t+1,x')\Big\}$$

gives us the desired property.

7. Consider a stochastic dynamic control problem where an agent has an exponential utility function, that is, they wish to minimize

$$J(t,x,U) = \mathbb{E}^U\Big[\exp\Big(\gamma\Big(\sum_{s=t}^{T} g(s,U_s) + \Phi\Big)\Big)\Big|\mathcal{F}_t\Big]$$

where $\gamma > 0$ is a risk aversion parameter. Assuming g is bounded, show that this problem admits a dynamic programming principle of the form

$$V_t = \operatorname*{ess\,inf}_{u\in\mathbb{U}} \mathbb{E}^U\Big[\exp\Big(\gamma g(t,U_t) + \log V_{t+1}\Big)\Big|\mathcal{F}_t\Big]$$

where $V_t = \operatorname{ess\,inf}_{U\in\mathbb{U}} J(t,U)$.

Sam Cohen: cohens@maths.ox.ac.uk

Solution: This is very similar to the construction given in lectures. We write

$$\begin{aligned}
J(t,U) &= \mathbb{E}^U\Big[\exp\Big(\gamma\Big(\sum_{s=t}^T g(s,U_s)+\Phi\Big)\Big)\Big|\mathcal{F}_t\Big]\\
&= \mathbb{E}^U\Big[e^{\gamma g(t,U_t)}\mathbb{E}^U\Big[\exp\Big(\gamma\Big(\sum_{s=t+1}^T g(s,U_s)+\Phi\Big)\Big)\Big|\mathcal{F}_{t+1}\Big]\Big|\mathcal{F}_t\Big]\\
&= \mathbb{E}^U\Big[e^{\gamma g(t,U_t)}J(t+1,U)\Big|\mathcal{F}_t\Big]
\end{aligned}$$

We optimize by taking $\tilde{U} = 1_{s\leq t}U + 1_{s>t}U'$, so, as our problem is dynamic, $J(t+1,\tilde{U}) = J(t+1,U')$ is independent of U. Hence, as our class of controls is closed under pasting,

$$\begin{aligned}
v(t,x) &= \operatorname*{ess\,inf}_{\tilde{U}} \mathbb{E}^{\tilde{U}}\Big[e^{\gamma g(t,\tilde{U}_t)}J(t+1,\tilde{U})\Big|\mathcal{F}_t\Big]\\
&= \operatorname*{ess\,inf}_{U,U'} \mathbb{E}^{U}\Big[e^{\gamma g(t,U_t)}J(t+1,U')\Big|\mathcal{F}_t\Big]\\
&= \operatorname*{ess\,inf}_{U} \mathbb{E}^{U}\Big[e^{\gamma g(t,U_t)}\operatorname*{ess\,inf}_{U'} J(t+1,U')\Big|\mathcal{F}_t\Big]\\
&= \operatorname*{ess\,inf}_{U} \mathbb{E}^{U}\Big[e^{\gamma g(t,U_t)}V_{t+1}\Big|\mathcal{F}_t\Big]
\end{aligned}$$

as desired.

Sam Cohen: cohens@maths.ox.ac.uk

8. Consider a model of a firm considering whether to comply with regulations (for example, deciding whether to emit pollution into a river system). At each time, the firm may be inspected, and will be penalized if they are found to be polluting. The inspections are random, but the regulator does not treat all firms the same – they split firms into two classes $\mathcal{P}$ (polluters), and $\mathcal{E}$ (environmental), and can choose the probabilities of inspection, and the penalties for pollution, separately for each set. The firms know what group they are in, and seek to optimize their expected discounted reward.

   Suppose that a firm in class $i \in \{\mathcal{E}, \mathcal{P}\}$ is inspected with probability ϕ_i. If found to violate it will be moved into $\mathcal{P}$ for the next period. If found not to be violating it will be moved into $\mathcal{E}$ for the next period. If a firm is inspected and is found to be violating, they will have to pay a fine F_i (depending on group). If a firm chooses not to violate, they pay a fixed cost c (in both groups). If a firm is not inspected, it remains in its current group.

   A firm facing this system must choose, at each time, the probability $u \in [0, 1]$ that it violates (it can randomize its choice, to avoid detection). Assume it wishes to minimize its cost, without considering other (environmental) impacts.

   (a) Write down the dynamic programming equation describing the value function of the firm.

   (b) Show that there is an optimal strategy for the firm which will simply violate, or not, in each state (that is, randomization of the strategy is unnecessary).

   (c) Assuming the firm acts over an infinite horizon, with discount rate $\rho > 0$, and $\phi_{\mathcal{P}}, \phi_{\mathcal{E}} \in (0, 1)$, calculate the cost-to-go function in each state for each unrandomized policy, and hence the value function.

   (d) Show that, if $\phi_{\mathcal{E}} F_{\mathcal{E}} \le \phi_{\mathcal{P}} F_{\mathcal{P}}$ (that is, past polluters are fined more heavily than past environmental firms), then it is never optimal for the firm to follow the rule 'violate when in $\mathcal{P}$ and comply when in $\mathcal{E}$'.

   Hint: for each policy U and state i, write $(1 - e^{-\rho})J(i, U)$ as a weighted average of payoffs, and then compare different policies.

   (e) Explain why, if

   $$\phi_{\mathcal{E}} F_{\mathcal{E}} < c < \phi_{\mathcal{P}} F_{\mathcal{P}},$$

   but $\phi_{\mathcal{P}} \ll \phi_{\mathcal{E}}$, the regulator will only be seen to issue small fines $F_{\mathcal{E}}$, and yet expected-profit-maximizing firms will usually comply with the environmental regulations.

   (This model is based on ideas due to Harrington (1988), *Enforcement leverage when penalties are restricted*, Journal of Public Economics 37:1, `https://doi.org/10.1016/0047-2727(88)90003-5`)

Sam Cohen: cohens@maths.ox.ac.uk

Solution:

(a) We write down the value in each state, knowing that there exists an optimal feedback policy. The state is the group that the firm is currently in. If in state i, the firm violates with probability u, and their expected fine is $\phi_i F_i u$, while their expected compliance cost is $(1-u)c$. The DPP then states

$$v(t,i) = \min_{u\in[0,1]} \Big\{ \underbrace{\phi_i F_i u}_{\text{average fine}} + \underbrace{(1-u)c}_{\text{compliance cost}} + e^{-\rho}\Big(\underbrace{\phi_i u v(t+1,\mathcal{P}) + \phi_i(1-u)v(t+1,\mathcal{E})}_{\text{if inspected}} + \underbrace{(1-\phi_i)v(t+1,i)}_{\text{not inspected}} \Big) \Big\}$$

(b) We observe that the minimization above is an affine function in u. Therefore either the firm is indifferent about whether to comply or not, or will choose $u \in \{0,1\}$ (so either way, no randomization is needed).

(c) Over an infinite horizon, we don't need dependence on t in the value function. Then our system of equations for the cost-to go, for $i \in \{\mathcal{E},\mathcal{P}\}$, is

$$J(i,U) = \phi_i F_i u_i + (1-u_i)c + e^{-\rho}\Big(\phi_i u_i J(\mathcal{P},U) + \phi_i(1-u_i)J(\mathcal{E},U) + (1-\phi_i)J(i,U)\Big)$$

We can then evaluate all four possible policies: $(u_{\mathcal{E}}, u_{\mathcal{P}}) \in \{(0,0),(0,1),(1,0),(1,1)\}$ (as in Question 2, for the cost to go). We write $J(\cdot,(ij)) = [J(\mathcal{E},(ij)), J(\mathcal{P},(ij))]^\top$ for the vector of costs under policy (i,j), so that we have the matrix-vector equa-

Sam Cohen: cohens@maths.ox.ac.uk

tions

$$
\begin{aligned}
J(\cdot,(00)) &= \begin{bmatrix} c \\ c \end{bmatrix} + e^{-\rho}\begin{bmatrix} 1 & 0 \\ \phi_{\mathcal{P}} & 1-\phi_{\mathcal{P}} \end{bmatrix} J(\cdot,(00)) = \frac{1}{1-e^{-\rho}}\begin{bmatrix} c \\ c \end{bmatrix}\\
J(\cdot,(01)) &= \begin{bmatrix} c \\ \phi_{\mathcal{P}}F_{\mathcal{P}} \end{bmatrix} + e^{-\rho}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} J(\cdot,(01)) = \frac{1}{1-e^{-\rho}}\begin{bmatrix} c \\ \phi_{\mathcal{P}}F_{\mathcal{P}} \end{bmatrix}\\
J(\cdot,(10)) &= \begin{bmatrix} \phi_{\mathcal{E}}F_{\mathcal{E}} \\ c \end{bmatrix} + e^{-\rho}\begin{bmatrix} 1-\phi_{\mathcal{E}} & \phi_{\mathcal{E}} \\ \phi_{\mathcal{P}} & 1-\phi_{\mathcal{P}} \end{bmatrix} J(\cdot,(10))\\
&= \begin{bmatrix} 1-e^{-\rho}(1-\phi_{\mathcal{E}}) & -e^{-\rho}\phi_{\mathcal{E}} \\ -e^{-\rho}\phi_{\mathcal{P}} & 1-e^{-\rho}(1-\phi_{\mathcal{P}}) \end{bmatrix}^{-1}\begin{bmatrix} \phi_{\mathcal{E}}F_{\mathcal{E}} \\ c \end{bmatrix}\\
&= \frac{1}{1-e^{-\rho}}\cdot\frac{1}{1+e^{-\rho}(\phi_{\mathcal{E}}+\phi_{\mathcal{P}}-1)}\begin{bmatrix} (1-e^{-\rho}(1-\phi_{\mathcal{P}}))\phi_{\mathcal{E}}F_{\mathcal{E}} + e^{-\rho}\phi_{\mathcal{E}}c \\ e^{-\rho}\phi_{\mathcal{P}}\phi_{\mathcal{E}}F_{\mathcal{E}} + (1-e^{-\rho}(1-\phi_{\mathcal{E}}))c \end{bmatrix}\\
&= \frac{1}{1-e^{-\rho}}\begin{bmatrix} \frac{1-e^{-\rho}(1-\phi_{\mathcal{P}})}{1-e^{-\rho}(1-\phi_{\mathcal{P}})+e^{-\rho}\phi_{\mathcal{E}}}\phi_{\mathcal{E}}F_{\mathcal{E}} + \frac{e^{-\rho}\phi_{\mathcal{E}}}{1-e^{-\rho}(1-\phi_{\mathcal{P}})+e^{-\rho}\phi_{\mathcal{E}}}c \\ \frac{e^{-\rho}\phi_{\mathcal{P}}}{1-e^{-\rho}(1-\phi_{\mathcal{E}})+e^{-\rho}\phi_{\mathcal{P}}}\phi_{\mathcal{E}}F_{\mathcal{E}} + \frac{1-e^{-\rho}(1-\phi_{\mathcal{E}})}{1-e^{-\rho}(1-\phi_{\mathcal{E}})+e^{-\rho}\phi_{\mathcal{P}}}c \end{bmatrix}\\
J(\cdot,(11)) &= \begin{bmatrix} \phi_{\mathcal{E}}F_{\mathcal{E}} \\ \phi_{\mathcal{P}}F_{\mathcal{P}} \end{bmatrix} + e^{-\rho}\begin{bmatrix} 1-\phi_{\mathcal{E}} & \phi_{\mathcal{E}} \\ 0 & 1 \end{bmatrix} J(\cdot,(11))\\
&= \frac{1}{1-e^{-\rho}}\cdot\begin{bmatrix} \frac{1-e^{-\rho}}{1-e^{-\rho}+e^{-\rho}\phi_{\mathcal{E}}}\phi_{\mathcal{E}}F_{\mathcal{E}} + \frac{e^{-\rho}\phi_{\mathcal{E}}}{1-e^{-\rho}+e^{-\rho}\phi_{\mathcal{E}}}\phi_{\mathcal{P}}F_{\mathcal{P}} \\ \phi_{\mathcal{P}}F_{\mathcal{P}} \end{bmatrix}
\end{aligned}
$$

The value function is simply the pointwise minimum of these four vectors (which will be achieved by one of the four, by the dynamic programming principle). Notice that, apart from the scaling term, the cost is always a weighted average of the relevant expected fine and compliance costs.

(d) In order for policy (01) to be optimal, we need $\phi_{\mathcal{P}}F_{\mathcal{P}} \le c$ (comparing with (00) in state $\mathcal{P}$). We know by assumption that $\phi_{\mathcal{E}}F_{\mathcal{E}} \le \phi_{\mathcal{P}}F_{\mathcal{P}}$, so comparing policy (01) with either policy (10) or (11) in either state, we see that (01) cannot be minimal (this is clear because of the weighted-average form of the cost-to-go functions.)

(e) If $\phi_{\mathcal{E}}F_{\mathcal{E}} < c < \phi_{\mathcal{P}}F_{\mathcal{P}}$ then we know that policy (00) is not optimal, by comparing with policy (10) in state $\mathcal{E}$. We also compare policy (10) with policy (11) in state $\mathcal{P}$, to see that (10) is preferable to (11). (In fact, we can see that this condition is then necessary and sufficient for (10) to be optimal). Therefore, profit maximizing firms will follow the strategy of 'violate when in state $\mathcal{E}$, comply when in state $\mathcal{P}$'. Under this policy, we have a Markov chain with transition matrix

$$
\begin{bmatrix} 1-\phi_{\mathcal{E}} & \phi_{\mathcal{E}} \\ \phi_{\mathcal{P}} & 1-\phi_{\mathcal{P}} \end{bmatrix}
$$

Sam Cohen: cohens@maths.ox.ac.uk

which has stationary distribution (i.e. left-eigenvector) given by

$$[\pi_{\mathcal{E}}, \pi_{\mathcal{P}}] = [\frac{\phi_{\mathcal{P}}}{\phi_{\mathcal{E}} + \phi_{\mathcal{P}}}, \frac{\phi_{\mathcal{E}}}{\phi_{\mathcal{E}} + \phi_{\mathcal{P}}}]$$

In particular, if $\phi_{\mathcal{P}}$ is small relative to $\phi_{\mathcal{E}}$, then firms will spend most of their time in state $\mathcal{P}$, in which they comply with the regulations. This leads to the curious situation where the regulator may issue no fines (so $F_{\mathcal{E}} = 0$, but inspect past polluters much less frequently than past compliers, and simply have the threat of issuing large fines without ever needing to do so).

Sam Cohen: cohens@maths.ox.ac.uk