ORF 418 - Optimal Learning - Fall 2024
Instructor: Emma Hubert

Midterm – Solution

October 23, in class (8:30am – 9:50am)

Exercise 1. A three-armed bandit problem. **[30pt]**

We consider a multi-armed bandit problem with **three arms**, $a \in \{1, 2, 3\}$. The random rewards from arm $a \in \{1, 2, 3\}$ are distributed according to a **Bernoulli** distribution with probability of success $a/4$. All the rewards are assumed to be independent.

We first **explore each arm** N **times** and obtain the rewards $\{r_k^{(a)}\}_{k=1,\dots N}$. We use the standard estimate

$$\widehat{q}(a) := \frac{1}{N} \sum_{k=1}^{N} r_k^{(a)}, \qquad a \in \{1, 2, 3\}.$$

and compute $a^\star := \arg\max_{a\in\{1,2,3\}} \widehat{q}(a)$. We further assume that, if there are more than one maximisers, we choose the **smallest**. We then **exploit** by using **only** $a^\star$ and receive independent rewards $(r_1, r_2, \dots)$. The limiting reward is defined as

$$J(a^\star) := \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} r_k.$$

a. **[5pt]** Assume $N = 1$ and $(r_1^{(1)}, r_1^{(2)}, r_1^{(3)}) = (0, 1, 1)$. Compute $a^\star$ and $J(a^\star)$.

Solution. If $(r_1^{(1)}, r_1^{(2)}, r_1^{(3)}) = (0, 1, 1)$, then $(\widehat{q}(1), \widehat{q}(2), \widehat{q}(3)) = (0, 1, 1)$, meaning that both arms 2 and 3 give the highest estimated expected reward. By assumption, we choose the 'smallest' arm between the two, *i.e.* $a^\star = 2$. We thus only use the second arm for the exploitation part. Therefore, the rewards $(r_k)_{k\in\mathbb{N}^\star}$ are i.i.d. and distributed according to a Bernoulli with probability of success $2/4 = 1/2$. By the law of large number, $J(a^\star)$ corresponds to the expected value of the rewards from the second arm, *i.e.*

$$J(a^\star) = \mathbb{E}\big[\mathrm{Ber}(1/2)\big] = 1/2.$$

b. **[5pt]** Compute $J(a^\star)$ for $a^\star = 1$ and $a^\star = 3$.

Solution. Let $a^\star = 1$, meaning that we will only use the first arm for the exploitation part. Therefore, the rewards $(r_k)_{k\in\mathbb{N}^\star}$ are i.i.d. and distributed according to a Bernoulli with probability of success $1/4$. By the law of large number, $J(a^\star)$ corresponds to the expected value of the rewards from the first arm, *i.e.*

$$J(a^\star) = \mathbb{E}\big[\mathrm{Ber}(1/4)\big] = 1/4.$$

Similarly, for $a^\star = 3$, we obtain

$$J(a^\star) = \mathbb{E}\big[\mathrm{Ber}(3/4)\big] = 3/4.$$

c. **[10pt]** Still assuming that $N = 1$, compute the distribution of $a^\star$.

Solution. Depending on the rewards $(r_1^{(1)}, r_1^{(2)}, r_1^{(3)})$, we could have either $a^\star = 1, 2$ or 3. More precisely:

$$a^\star = \begin{cases} 1 & \text{if } r_1^{(1)} \geq r_1^{(2)} \text{ and } r_1^{(1)} \geq r_1^{(3)} \\ 2 & \text{if } r_1^{(2)} > r_1^{(1)} \text{ and } r_1^{(2)} \geq r_1^{(3)} \\ 3 & \text{if } r_1^{(3)} > r_1^{(1)} \text{ and } r_1^{(3)} > r_1^{(2)}. \end{cases}$$

Recalling that the rewards $(r_1^{(1)}, r_1^{(2)}, r_1^{(3)})$ are discrete random variables taking values in $\{0,1\}$, we need to distinguish different cases. For example, the inequality $r_1^{(1)} \geq r_1^{(2)}$ would be true if $(r_1^{(1)}, r_1^{(2)}) = (1,0)$ but also if $(r_1^{(1)}, r_1^{(2)}) = (1,1)$ or if $(r_1^{(1)}, r_1^{(2)}) = (0,0)$. It is easier to start by computing $\mathbb{P}(a^\star = 3)$, as the inequalities $r_1^{(3)} > r_1^{(1)}$ and $r_1^{(3)} > r_1^{(2)}$ are both true if and only if $(r_1^{(1)}, r_1^{(2)}, r_1^{(3)}) = (0,0,1)$. By independance of the rewards, we thus have:

$$\begin{aligned}\mathbb{P}(a^\star = 3) &= \mathbb{P}\big(r_1^{(3)} > r_1^{(1)}, r_1^{(3)} > r_1^{(2)}\big) = \mathbb{P}\big(r_1^{(1)} = 0, r_1^{(2)} = 0, r_1^{(3)} = 1\big) = \mathbb{P}\big(r_1^{(1)} = 0\big)\mathbb{P}\big(r_1^{(2)} = 0\big)\mathbb{P}\big(r_1^{(3)} = 1\big) \\ &= \frac{3}{4} \times \frac{1}{2} \times \frac{3}{4} = \frac{9}{32}.\end{aligned}$$

To compute $\mathbb{P}(a^\star = 2)$, we can notice that having $r_1^{(2)} > r_1^{(1)}$ necessarily requires $(r_1^{(1)}, r_1^{(2)}) = (0,1)$, which implies $r_1^{(2)} \geq r_1^{(3)}$, regardless of the value of $r_1^{(3)}$. Therefore

$$\mathbb{P}(a^\star = 2) = \mathbb{P}\big(r_1^{(2)} > r_1^{(1)}, r_1^{(2)} \geq r_1^{(3)}\big) = \mathbb{P}\big(r_1^{(1)} = 0, r_1^{(2)} = 1\big) = \mathbb{P}\big(r_1^{(1)} = 0\big)\mathbb{P}\big(r_1^{(2)} = 1\big) = \frac{3}{4} \times \frac{1}{2} = \frac{3}{8}.$$

Finally, we can easily compute

$$\mathbb{P}(a^\star = 1) = 1 - \mathbb{P}(a^\star = 2) - \mathbb{P}(a^\star = 3) = \frac{32}{32} - \frac{12}{32} - \frac{9}{32} = \frac{11}{32}.$$

Therefore, the distribution of $a^\star$ is given by

$$a^\star = \begin{cases} 1 & \textbf{with probability } 11/32 \\ 2 & \textbf{with probability } 12/32 \\ 3 & \textbf{with probability } 9/32. \end{cases}$$

d. [10pt] We now assume $N \longrightarrow +\infty$. Deduce $a^\star$ and compute the limiting reward if, instead of using only $a^\star$, we implement an ε-greedy algorithm with $\varepsilon = 0.2$ (briefly describe the algorithm).

Solution. For $N \longrightarrow +\infty$, by the law of large numbers, we have for all $a \in \{1,2,3\}$,

$$\lim_{N\to\infty} \widehat{q}(a) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} r_k^{(a)} = \mathbb{E}\big[\mathrm{Ber}(a/4)\big] = \frac{a}{4}.$$

It is thus clear that, since arm 3 gives the best expected reward, $a^\star = 3$. We now implement a ε-greedy algorithm with $\varepsilon = 0.2$, meaning that at each time step $k \in \mathbb{N}^\star$, we choose

$$A_k = \begin{cases} 1 & \textbf{with probability } 0.1 \\ 2 & \textbf{with probability } 0.1 \\ 3 & \textbf{with probability } 0.8. \end{cases}$$

Using the results from questions a. and b., we then obtain

$$\mathbb{E}[J(A)] = J(1) \times 0.1 + J(2) \times 0.1 + J(3) \times 0.8 = \frac{1}{4} \times \frac{1}{10} + \frac{2}{4} \times \frac{1}{10} + \frac{3}{4} \times \frac{8}{10} = \frac{27}{40}.$$

Exercise 2. A (linear quadratic?) control problem. **[25pt]**

Consider the following one-dimensional control problem with state dynamics

$$x_{k+2} - 0.5x_k = bu_k,\ k \in \mathbb{N},$$

for $(x_0, x_1) \in \mathbb{R}^2$ and $b \in \mathbb{R}$. The cost function (to be minimised) is given by

$$J(x_0, x_1, u) := \sum_{k=0}^{\infty} \left(x_k^2 + u_k^2\right).$$

a. **[5pt]** Show by mathematical induction that, if $b = 0$, then $x_{2k} = 0.5^k x_0$ and $x_{2k+1} = 0.5^k x_1$, for all $k \in \mathbb{N}$.

Solution. We can easily check that for $k = 0$, the previous formula give exactly x_0 and x_1. Assume now that the formula are true for some $k \in \mathbb{N}$, we want to show that there are still true for $k + 1$, meaning

$$x_{2(k+1)} = 0.5^{k+1} x_0, \quad \textbf{and} \quad x_{2(k+1)+1} = 0.5^{k+1} x_1.$$

We can first compute, using the dynamics for $b = 0$ and then the induction hypothesis, that

$$x_{2k+2} = 0.5 x_{2k} = 0.5 \times 0.5^k x_0 = 0.5^{k+1} x_0.$$

Similarly,

$$x_{2k+3} = 0.5 x_{2k+1} = 0.5 \times 0.5^k x_1 = 0.5^{k+1} x_1.$$

b. **[5pt]** Formulate the previous control problem as a standard linear-quadratic problem in dimension $d = 2$ (write A, B, M, N and ρ). When is the system controllable?

Solution. To formulate the previous control problem as a standard linear-quadratic problem in dimension $d = 2$, we define a new state $y_k := (x_k, x_{k+1})^\top$ starting from $y_0 := (x_0, x_1)^\top$. We can then write:

$$y_{k+1} = \begin{pmatrix} x_{k+1} \\ x_{k+2} \end{pmatrix} = A \begin{pmatrix} x_k \\ x_{k+1} \end{pmatrix} + Bu_k, \quad \textbf{for} \quad A := \begin{pmatrix} 0 & 1 \\ 0.5 & 0 \end{pmatrix} \quad \textbf{and} \quad B := \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

Then, take $N := 1$ and

$$M := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

so that the cost $J(x_0, x_1, u)$ can be rewritten as

$$J(y_0, u) = \sum_{k=0}^{\infty} \left(y_k^\top M y_k + u_k^\top N u_k\right).$$

Given the previous parameters, the controllability matrix is defined by:

$$\mathcal{C} := \begin{pmatrix} B & AB \end{pmatrix} = \begin{pmatrix} 0 & b \\ b & 0 \end{pmatrix}.$$

The previous matrix has full rank 2 if and only if $b \neq 0$. Therefore, if $b \neq 0$ the system is controllable, otherwise it is not.

c. **[5pt]** Define the value function v corresponding to this control problem. Is the value function finite?

Solution. The value function for this infinite horizon deterministic LQ control problem is defined as usual by:

$$v(y_0) := \inf_u J(y_0, u), \quad y_0 := (x_0, x_1) \in \mathbb{R}^2.$$

For $b \neq 0$, we proved in the previous question that the system is controllable. Therefore, if $b \neq 0$, the value function is finite. It remains to study the case when $b = 0$. In question a., we proved that if $b = 0$, we have $x_{2k} = 0.5^k x_0$ and $x_{2k+1} = 0.5^k x_1$, for all $k \in \mathbb{N}$, and in particular, the system is not controlled. Choosing $u = (0, 0, \dots)$, we can thus rewrite the cost as:

$$J(x_0, x_1, u) = \sum_{i=0}^{\infty} x_{2i}^2 + \sum_{i=0}^{\infty} x_{2i+1}^2 + \sum_{k=0}^{\infty} u_k^2 \leq \sum_{i=0}^{\infty} \left(0.5^i x_0\right)^2 + \sum_{i=0}^{\infty} \left(0.5^i x_1\right)^2 = \left(x_0^2 + x_1^2\right) \sum_{i=0}^{\infty} 0.5^{2i}.$$

Since $0 < 0.5^2 < 1$, the sum is finite, implying that the cost, and therefore also the value function, are finite.

d. **[10pt]** State the general Dynamic Programming Equation and use it to show that the value function satisfies:

$$v(x_0, x_1) = x_0^2 + \inf_{u \in \mathbb{R}} \left\{u^2 + v(\widetilde{x})\right\}, \qquad \forall (x_0, x_1) \in \mathbb{R}^2, \tag{2.1}$$

for $\widetilde{x} \in \mathbb{R}^2$ to be determined ($\widetilde{x}$ is allowed to depend on x_0, x_1, u and the parameters of the model).

Solution. The general Dynamic Programming equation (Thm 2.2.1 of the Lecture notes) is:

$$v(y) = y^\top M y + \inf_{u \in \mathbb{R}^\ell} \left\{u^\top N u + \rho v(Ay + Bu)\right\}, \quad \textbf{for all} \quad y \in \mathbb{R}^2.$$

Using the parameters from question b., namely

$$A := \begin{pmatrix} 0 & 1 \\ 0.5 & 0 \end{pmatrix} \; B := \begin{pmatrix} 0 \\ b \end{pmatrix} \; M := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \; N := 1, \quad \textbf{and} \quad \rho = 1,$$

we get:

$$v(x_0, x_1) = x_0^2 + \inf_{u \in \mathbb{R}} \left\{u^2 + v\big(x_1, 0.5x_0 + bu\big)\right\}, \qquad \forall (x_0, x_1) \in \mathbb{R}^2.$$

Therefore, $\widetilde{x} := (x_1, 0.5x_0 + bu)$.

Exercise 3. Bayesian estimation. **[25pt]**

Let $\theta \in (0,1)$ be the **unknown** proportion of individuals infected by a virus in a given population. To estimate θ, we test 100 individuals and denote by Y the number of infected individuals among them. We observe $Y = 10$.

Hint. The Beta distribution with parameters $a, b > 0$ is characterised by a p.d.f. (density) given by

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}\,(1-x)^{b-1}, \qquad x \in (0,1),$$

where Γ is the gamma function, which satisfies in particular $\Gamma(1) = 1$ and $\Gamma(x+1) = x\Gamma(x)$. Moreover, its mean is equal to $a/(a+b)$.

---

a. **[10pt]** Write the conditional probability $\mathbb{P}(Y = y|\theta)$ for any $y \in \{0, \dots, 100\}$ and compute the maximum likelihood estimator corresponding to the observation $Y = 10$.

Solution. Let $Y = Y_1 + \ldots + Y_{100}$. Knowing $\theta \in (0,1)$, the r.v. Y_k are i.i.d Bernoulli with success probability θ. Therefore, $Y|\theta$ is a Binomial r.v. with parameters $(100, \theta)$. We thus have:

$$\mathbb{P}(Y = y|\theta) = \binom{100}{y} \theta^y (1-\theta)^{100-y}.$$

To compute the maximum likelihood estimator corresponding to the observation $Y = 10$, it is more convenient to study here the log-likelihood:

$$\mathcal{L}(\theta) := \ln\big(\mathbb{P}(Y = 10|\theta)\big),\ \theta \in (0,1).$$

Using $\mathbb{P}(Y = y|\theta)$ written above for $y = 10$, we obtain:

$$\mathcal{L}(\theta) = \ln\big(\mathbb{P}(Y = 10|\theta)\big) = \ln\left(\binom{100}{10}\theta^{10}(1-\theta)^{90}\right) = \ln\left(\binom{100}{10}\right) + 10\ln(\theta) + 90\ln(1-\theta).$$

The FOC condition gives

$$\frac{10}{\theta} - \frac{90}{1-\theta} = 0, \qquad \textbf{\textit{i.e.}} \qquad \theta = \frac{1}{10}.$$

Note that this estimator simply corresponds to the proportion of infected individual among the tested individuals.

b. **[10pt]** We now assume that the prior distribution for θ is a Beta distribution with parameters $a > 0$ and $b > 0$. Show that the posterior distribution for θ given the observation $Y = 10$ is again a Beta distribution, but with parameters $(a + 10, b + 90)$.

Solution. Computing the posterior distribution for θ given the observation $Y = 10$ corresponds to computing the density of $\theta|Y = 10$. To do so, we can use the following Bayes formula,

$$f_{\theta|Y=10}(x) = \frac{\mathbb{P}(Y = 10|\theta = x) f_\theta(x)}{\mathbb{P}(Y = 10)},$$

with

$$f_\theta(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{x\in(0,1)},$$

$$\mathbb{P}(Y = 10|\theta = x) = \binom{100}{10} x^{10}(1-x)^{90},$$

$$\textbf{and}\ \ \mathbb{P}(Y = 10) = \int \mathbb{P}(Y = 10|\theta = x) f_\theta(x)\mathrm{d}x = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\binom{100}{10}\int_0^1 x^{a+10-1}(1-x)^{b+90-1}\mathrm{d}x.$$

We thus obtain

$$f_{\theta|Y=10}(x) = C^{-1}x^{a+10-1}(1-x)^{b+90-1}\mathbb{1}_{x\in(0,1)}, \quad \textbf{with} \quad C := \int_0^1 x^{a+10-1}(1-x)^{b+90-1}\mathrm{d}x.$$

Recall that the density of a Beta distribution with parameters (α, β) should integrate to 1, meaning that

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = 1, \quad \textbf{\textit{i.e.}} \quad \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Using this for $\alpha := a+10$ and $\beta := b+90$, we thus obtain

$$C := \frac{\Gamma(a+10)\Gamma(b+90)}{\Gamma(a+b+100)}.$$

We thus recognize that the posterior density

$$f_{\theta|Y=10}(x) = \frac{\Gamma(a+10)\Gamma(b+90)}{\Gamma(a+b+100)}x^{a+10-1}(1-x)^{b+90-1}\mathbb{1}_{x\in(0,1)},$$

coincides with the density of a Beta distribution with parameters $\alpha := a+10$ and $\beta := b+90$.

c. **[5pt]** Compute the Bayesian estimator (for a quadratic loss) and compare it to the MLE estimator. Comment.

Solution. We previously obtained that $\theta|Y = 10$ follows a Beta distribution with parameters $\alpha := a+10$ and $\beta := b+90$. Therefore, the Bayesian estimator for θ is given by

$$\widehat{\theta} = \mathbb{E}[\theta|Y=10] = \frac{\alpha}{\alpha+\beta} = \frac{a+10}{a+b+100}.$$

From question a., we know that the MLE estimator is $\theta = 1/10$. We thus have $\widehat{\theta} > 1/10$ if and only if $9a > b$. Otherwise, if $9a < b$, then $\widehat{\theta} < 1/10$. In particular, if $9a = b$, then both estimators coincide.

Exercise 4. Another control problem. **[20pt]**

Consider the following one-dimensional ($d = \ell = 1$) control problem with finite time horizon $n \in \mathbb{N}^\star$ and state dynamics

$$x_{k+1} = x_k + bu_k, \quad k \in \{0, \dots, n-1\},$$

for some $b \in \mathbb{R}$ and $x_0 \in \mathbb{R}$ fixed. We denote by $\bar{u}_k := (u_k, \dots, u_{n-1})$ the sequence of controls starting from step $k \in \{0, \dots, n-1\}$. The **reward** function when starting at step $k \in \{0, \dots, n\}$ from state $x \in \mathbb{R}$ is defined by

$$J\big(k, x, \bar{u}_k\big) := \rho^{n-k} x_n - \sum_{i=k}^{n-1} \rho^{i-k} u_i^2, \ k \in \{0, \dots, n-1\}, \text{ and } J(n, x) = x,$$

for some **discount factor** $\rho \in (0, 1)$. Finally, we define the value function of this **maximisation** problem, when starting from state $x \in \mathbb{R}$ at step $k \in \{0, \dots, n-1\}$, by

$$v(k, x) := \sup_{\bar{u}_k} J(k, x, \bar{u}_k).$$

a. **[5pt]** Give a sufficient and necessary condition for the system to be controllable.

Solution. Even if this control problem does not completely correspond to a linear-quadratic control problem, the dynamic is linear so we can apply the usual result on controllability. More precisely, in dimension one, we can easily compute that the controllability matrix $\mathcal{C} := (b)$ has full rank if and only if $b \neq 0$. Therefore, the system is controllable if and only if $b \neq 0$. Note that you can also use the definition of controllability to show that the system is controllable if and only if $b \neq 0$.

b. **[5pt]** Show that for all $k \in \{0, \dots, n-1\}$,

$$J\big(k, x, \bar{u}_k\big) = -u_k^2 + \rho J\big(k+1, x + bu_k, \bar{u}_{k+1}\big).$$

Solution. Using the definition of the cost, we first separate the cost at step k from the rest:

$$J\big(k, x, \bar{u}_k\big) := \rho^{n-k} x_n - \sum_{i=k}^{n-1} \rho^{i-k} u_i^2 = -\rho^0 u_k^2 + \rho^{n-k} x_n - \sum_{i=k+1}^{n-1} \rho^{i-k} u_i^2.$$

To make the cost $J\big(k+1, x + bu_k, \bar{u}_{k+1}\big)$ appear, remark that:

$$\rho^{n-k} x_n - \sum_{i=k+1}^{n-1} \rho^{i-k} u_i^2 = \rho\left(\rho^{n-(k+1)} x_n - \sum_{i=k+1}^{n-1} \rho^{i-(k+1)} u_i^2\right) = \rho J\big(k+1, x + bu_k, \bar{u}_{k+1}\big).$$

We therefore conclude as requested that

$$J\big(k, x, \bar{u}_k\big) = -u_k^2 + \rho J\big(k+1, x + bu_k, \bar{u}_{k+1}\big).$$

This was not required to get full point at the question, but if you wanted to prove that the state being considered at step $k+1$ should be $x + bu_k$, you can prove by mathematical induction that

$$x_n = x_k + b \sum_{i=k}^{n-1} u_i, \qquad \textbf{for } \ k \in \{0, \dots, n-1\}.$$

This corresponds to the final value (at step n) of the state when starting from x_k at step k and using the sequence of controls $\bar{u}_k := (u_k, \dots, u_{n-1})$ Therefore, if starting at step $k \in \{0, \dots, n-1\}$ from state $x \in \mathbb{R}$, we can write

$$x_n = x + b \sum_{i=k}^{n-1} u_i = x + b\left(u_k + \sum_{i=k+1}^{n-1} u_i\right) = \big(x + bu_k\big) + b \sum_{i=k+1}^{n-1} u_i,$$

which corresponds to x_n when starting from $x + bu_k$ at step $k+1$ and using the sequence of controls $\bar{u}_{k+1} := (u_{k+1}, \dots, u_{n-1})$.

c. **[10pt]** Deduce the dynamic programming equation satisfied by the value function $v(k,x)$ for $k \in \{0,\dots,n-1\}$ and $x \in \mathbb{R}$. Specify $v(n,x)$.

Solution. From the previous question, we have

$$J\big(k,x,\bar{u}_k\big) = -u_k^2 + \rho J\big(k+1, x+bu_k, \bar{u}_{k+1}\big).$$

To obtain an equation for the value function, we take the supremum over $\bar{u}_k$ on both sides:

$$v\big(k,x\big) := \sup_{\bar{u}_k} J\big(k,x,\bar{u}_k\big) = \sup_{\bar{u}_k} \big\{ -u_k^2 + \rho J\big(k+1, x+bu_k, \bar{u}_{k+1}\big)\big\}.$$

We then split the maximisation over $\bar{u}_k$ into a maximisation over u_k and then $\bar{u}_{k+1}$. This gives:

$$v(k,x) = \sup_{u_k} \big\{ -u_k^2 + \rho \sup_{\bar{u}_{k+1}} J\big(k+1, x+bu_k, \bar{u}_{k+1}\big)\big\}.$$

Noticing that

$$\sup_{\bar{u}_{k+1}} J\big(k+1, x+bu_k, \bar{u}_{k+1}\big) = v\big(k+1, x+bu_k\big),$$

we deduce the following equation for v:

$$v(k,x) = \sup_{u_k} \big\{ -u_k^2 + \rho v\big(k+1, x+bu_k\big)\big\}.$$

Finally, $v(n,x)$ is the value when starting at the finite time horizon. At that time, there is no control to be chosen, and thus $v(n,x) = x$.