Benoît Chachuat
ME C2 401, Ph: 33844, benoit.chachuat@epfl.ch

Problem Set #1 (With Corrections)

1. What is the optimal solution of an unconstrained minimization problem with linear cost function, $f(\mathbf{x}) = \mathbf{c}^{\mathsf{T}}\mathbf{x}$, with $\mathbf{c} \neq \mathbf{0}$?

Solution. Suppose that $\mathbf{c} \in \mathbb{R}^n$, and let $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that $\bar{x}_i = \operatorname{sign}(c_i) \times b$, $i = 1, \ldots, n$. Then,

$$\lim_{b \downarrow -\infty} \mathbf{c}^{\mathsf{T}}\bar{\mathbf{x}} \;=\; \left(\sum_{i=1}^{n} |c_i|\right) \lim_{b \downarrow -\infty} b \;=\; -\infty.$$

Therefore, there is no optimal solution to that problem.

---

2. Given a cardboard of area A to make a rectangular box, what is the maximum volume that can be attained?

Solution. Let x_1, x_2 and x_3 denote the edges of the box. Mathematically, the problem formulates as:

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & -x_1x_2x_3 \\ \text{s.t.} \quad & 2(x_1x_2 + x_1x_3 + x_2x_3) - A = 0 \\ & x_1, x_2, x_3 \geq 0 \end{aligned} \tag{1}$$

Note first that $x_1 = x_2 = 0$ cannot be a solution to the problem, for it does not satisfy the constraint. Now, a possible (exact) reformulation of this problem consists of expressing x_3 as a function of x_1 and x_2 based on the constraint,

$$x_3 \;=\; \frac{\frac{1}{2}A - x_1x_2}{x_1 + x_2}.$$

Substituting x_3 for this expression in Problem (1) yields the problem (in $\mathbb{R}^2$):

$$\begin{aligned} \min_{x_1,x_2} \; & f(\mathbf{x}) := -\frac{x_1x_2}{x_1+x_2}\left(\frac{1}{2}A - x_1x_2\right) \\ \text{s.t.} \quad & \mathbf{x} \in S := \{\mathbf{x} : x_1, x_2 \geq 0 \text{ and } x_1x_2 \leq A\}. \end{aligned} \tag{2}$$

We shall address this problem by ignoring the inequality constraints, then determine candidate unconstrained solutions, and finally check that these solutions belong to the feasible domain. The gradient of f is given by

$$\boldsymbol{\nabla} f(\mathbf{x}) \;=\; \begin{pmatrix} -\dfrac{x_1x_2}{x_1+x_2}\left[\dfrac{A}{2x_1} - 2x_2 + \dfrac{1}{x_1+x_2}\left(x_1x_2 - \dfrac{A}{2}\right)\right] \\ -\dfrac{x_1x_2}{x_1+x_2}\left[\dfrac{A}{2x_2} - 2x_1 + \dfrac{1}{x_1+x_2}\left(x_1x_2 - \dfrac{A}{2}\right)\right] \end{pmatrix}.$$

Since every (local) minimum point $\mathbf{x}^\star$ of f is a stationary point, $\boldsymbol{\nabla} f(\mathbf{x}^\star) = \mathbf{0}$. That is, the difference between the two gradient components of f is also zero, and we get

$$\frac{x_1x_2}{x_1+x_2}\left[\frac{A}{2}\left(\frac{1}{x_1}-\frac{1}{x_2}\right)+2(x_1-x_2)\right] = \frac{x_1x_2}{x_1+x_2}\,(x_1-x_2)\left(2-\frac{A}{2x_1x_2}\right) = 0.$$

We have one of three cases:

- If $x_1^\star = x_2^\star \neq 0$, then the stationarity conditions give $x_1^\star = x_2^\star = \sqrt{\frac{A}{6}}$. ($x_1^\star = x_2^\star = -\sqrt{\frac{A}{6}}$ is also a stationary point of f, but it lies outside the feasible domain.)
- If $x_1^\star x_2^\star = \frac{A}{4}$, then for $\mathbf{x}^\star$ to satisfy the stationarity conditions, we must have

$$\begin{aligned} 0 &= \frac{A}{2x_1^\star} - 2x_2^\star + \frac{1}{x_1^\star + x_2^\star}\left(x_1^\star x_2^\star - \frac{A}{2}\right) \\ &= -\frac{1}{x_1^\star + \frac{A}{4x_1^\star}}\frac{A}{4}, \end{aligned}$$

  which is clearly impossible.
- Finally, if either $x_1^\star = 0$ or $x_2^\star = 0$, then $f(\mathbf{x}^\star) = 0$, and $\mathbf{x}^\star$ cannot be a minimum for the problem since

$$f\left(\sqrt{\frac{A}{6}}, \sqrt{\frac{A}{6}}\right) = -\left(\frac{A}{6}\right)^{\frac{3}{2}} < 0.$$

  there is always a point with negative objective function value in any neighborhood $\mathbf{x}^\star$ (e.g., $x_1 = \varepsilon > 0$ as $\varepsilon \to 0$).

A representation of the objective function f is shown in Fig. 1, for $A = 1$. It is seen that the candidate stationary point $x_1^\star = x_2^\star = \sqrt{\frac{A}{6}}$ is indeed a local minimum point of f. The corresponding minimum point for the original problem (1) is

$$x_1^\star = x_2^\star = x_3^\star = \sqrt{\frac{A}{6}} \qquad \text{and} \qquad f\left(\mathbf{x}^\star\right) = -\left(\frac{A}{6}\right)^{\frac{3}{2}}.$$

---

3. Consider P points $(x_1, y_1), \ldots, (x_P, y_P) \in \mathbb{R}^2$. Given a set of basis functions, $b_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$, find the coefficients $p_1, \ldots, p_n$ in such a way that the curve

$$y = \sum_{i=1}^{n} p_i b_i(x),$$

gives the minimum least square error

$$f(\mathbf{p}) := \sum_{k=1}^{P}\left(y_k - \sum_{i=1}^{n} p_i b_i(x_k)\right)^2.$$

Solution. It shall be assumed throughout that $P > n$, i.e., there are more points than coefficients to estimate, and that the following $(n \times P)$ matrix is full rank:

$$\mathsf{B} := \begin{pmatrix} b_1(x_1) & \cdots & b_n(x_1) \\ \vdots & \ddots & \vdots \\ b_1(x_P) & \cdots & b_n(x_P) \end{pmatrix}.$$

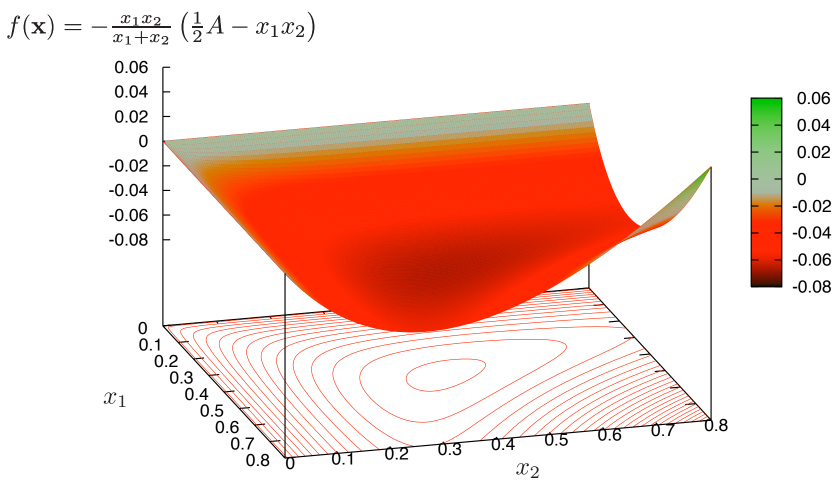

Figure 1: Illustration of the objective function in Exercise 2, for $A = 1$.

Candidate minimum points for the problem to minimize $f(\mathbf{p})$ for $\mathbf{p} \in \mathbb{R}^n$ are the stationary points of f, i.e., points satisfying the condition $\nabla_{p_i} f(\mathbf{p}) = 0$, for each $i = 1, \ldots, n$. Based on the definition of f, we have,

$$\begin{aligned}
\nabla_{p_i} f(\mathbf{p}) &= -2 \sum_{k=1}^{P} b_i(x_k) \left(y_k - \sum_{j=1}^{n} p_j b_j(\mathbf{x}_k) \right) \\
&= -2 \sum_{k=1}^{P} b_i(x_k) y_k + 2 \sum_{j=1}^{n} \left(p_j \sum_{k=1}^{P} b_i(\mathbf{x}_k) b_j(\mathbf{x}_k) \right).
\end{aligned}$$

Hence, the stationary condition reads

$$\sum_{j=1}^{n} \left(p_j^\star \sum_{k=1}^{P} b_i(\mathbf{x}_k) b_j(\mathbf{x}_k) \right) = \sum_{k=1}^{P} b_i(x_k) y_k \qquad i = 1, \ldots, n,$$

or, equivalently, in matrix notation,

$$\mathsf{B}^\mathsf{T}\mathsf{B}\mathbf{p}^\star = \mathsf{B}^\mathsf{T}\mathbf{y},$$

with $\mathbf{y} := \begin{pmatrix} y_1 & \cdots & y_P \end{pmatrix}^\mathsf{T}$. Finally, B being full rank, $\mathsf{B}^\mathsf{T}\mathsf{B}$ is invertible, and we get

$$\mathbf{p}^\star = \left(\mathsf{B}^\mathsf{T}\mathsf{B}\right)^{-1} \mathsf{B}^\mathsf{T}\mathbf{y}.$$

---

4. (*) In this problem, we consider a discrete-time dynamic system of the form

$$\begin{aligned}
x_{i+1} &= f_i(x_i, u_i) && (3)\\
x_0 &= a \text{ (given)}, && (4)
\end{aligned}$$

and try to find a control sequence $\mathbf{u} = (u_0, \ldots, u_{N-1})^\mathsf{T}$ that minimizes the function

$$\mathcal{J}(\mathbf{u}) = g_N(x_N) + \sum_{i=0}^{N-1} g_i(x_i, u_i).$$

In this objective, we wish to calculate the gradient of the objective function $\mathcal{J}$ with respect to the control variables $\mathbf{u}$.

I) Sensitivity-Based Approach for Gradient Calculation.

(a) Assume that a solution to the difference equation (3) exists and is unique, and can be written as an explicit function of the control variables,

$$x_i \;=\; \phi_i(u_0, \ldots, u_{i-1}).$$

Show that the gradient of the objective function $\mathcal{J}$ is given by

$$\nabla_{u_k}\mathcal{J} \;=\; \frac{\partial g_k}{\partial u_k} + \sum_{i=k+1}^{N} \frac{\partial g_i}{\partial x_i}\frac{\partial \phi_i}{\partial u_k}, \tag{5}$$

for each $k = 1, \ldots, N-1$.

Solution. Substituting each x_i by its explicit expression $\phi_i(u_0, \ldots, u_{i-1})$ in $\mathcal{J}$, then applying a chain rule of differentiation, gives

$$\nabla_{u_k}\mathcal{J} \;=\; \frac{\partial g_N}{\partial x_N}\frac{\partial \phi_N}{\partial u_k} + \sum_{i=0}^{N-1}\left(\frac{\partial g_i}{\partial x_i}\frac{\partial \phi_i}{\partial u_k} + \frac{\partial g_i}{\partial u_k}\right).$$

Observing that $\frac{\partial \phi_i}{\partial u_k} = 0$ for $k \geq i$, and $\frac{\partial g_i}{\partial u_k} = 0$ for $k \neq i$, we get

$$\nabla_{u_k}\mathcal{J} \;=\; \frac{\partial g_k}{\partial u_k} + \sum_{i=k+1}^{N} \frac{\partial g_i}{\partial x_i}\frac{\partial \phi_i}{\partial u_k}.$$

(b) Using a chain rule of differentiation, derive a recursive expression for the state sensitivities, i.e., a mathematical expression giving $\frac{\partial x_{i+1}}{\partial u_k}$ as a function of $\frac{\partial x_i}{\partial u_k}$. Also derive initial conditions for the state sensitivities $\frac{\partial x_0}{\partial u_k}$.

Solution. Applying a chain rule of differentiation to the difference equation (3) gives

$$\frac{\partial x_{i+1}}{\partial u_k} \;=\; \frac{\partial f_i}{\partial x_i}\frac{\partial x_i}{\partial u_k} + \frac{\partial f_i}{\partial u_k},$$

for each $k = 0, \ldots, N-1$ and each $i = 1, \ldots, N-1$. Regarding the state sensitivities at $k = 0$, since $x_0 = a$ is fixed, its value does not depend on any control $u_0, \ldots, u_{N-1}$, and we have

$$\frac{\partial x_0}{\partial u_k} \;=\; 0,$$

for each $k = 0, \ldots, N-1$.

(c) Evaluating the gradient $\boldsymbol{\nabla}_{\mathbf{u}}\mathcal{J}$ by directly substituting the state sensitivities $\frac{\partial x_i}{\partial u_k}$ in Eq. (5) is often referred to as the *(forward) sensitivity approach* in the literature. How many finite difference equations is it necessary to solve for evaluating the gradient of the objective function in this approach? And how can we expect the (computational) cost of a gradient evaluation to grow with the number N of control variables?

Solution. An estimate of the gradient $\nabla_{\mathbf{u}}\mathcal{J}$ is obtained as

$$\nabla_{\mathbf{u}}\mathcal{J} = \begin{pmatrix} \frac{\partial g_0}{\partial u_0} + \sum_{i=1}^{N} \frac{\partial g_i}{\partial x_i}\frac{\partial x_i}{\partial u_0} \\ \frac{\partial g_1}{\partial u_1} + \sum_{i=2}^{N} \frac{\partial g_i}{\partial x_i}\frac{\partial x_i}{\partial u_1} \\ \vdots \\ \frac{\partial g_{N-1}}{\partial u_{N-1}} + \frac{\partial g_N}{\partial x_N}\frac{\partial x_N}{\partial u_{N-1}} \\ \frac{\partial g_N}{\partial u_N} \end{pmatrix},$$

and requires that x_i and $\frac{\partial x_i}{\partial u_0}, \ldots, \frac{\partial x_i}{\partial u_{N-1}}$ be calculated, at each $i = 0, \ldots, N$, by solving the difference equations

$$\begin{aligned} x_{i+1} &= f_i(x_i, u_i) \\ \frac{\partial x_{i+1}}{\partial u_0} &= \frac{\partial f_i}{\partial x_i}\frac{\partial x_i}{\partial u_0} + \frac{\partial f_i}{\partial u_0} \\ &\vdots \\ \frac{\partial x_{i+1}}{\partial u_{N-1}} &= \frac{\partial f_i}{\partial x_i}\frac{\partial x_i}{\partial u_{N-1}} + \frac{\partial f_i}{\partial u_{N-1}}, \end{aligned}$$

from the initial conditions,

$$\begin{aligned} x_0 &= a \\ \frac{\partial x_{i+1}}{\partial u_0} &= 0 \\ &\vdots \\ \frac{\partial x_{i+1}}{\partial u_{N-1}} &= 0. \end{aligned}$$

Overall, each evaluation of $\nabla_{\mathbf{u}}\mathcal{J}$ thus requires that $N + 1$ difference equations be integrated forward in time. That is, the increase in the computational cost of a gradient evaluation is roughly *linear* in the number N of control variables.

II) Adjoint-Based Approach for Gradient Calculation.

(a) Consider the augmented function

$$\widehat{\mathcal{J}} := \mathcal{J} + \sum_{i=0}^{N-1} p_{i+1}\left[f_i(x_i, u_i) - x_{i+1}\right],$$

which is obviously equal to $\mathcal{J}$, as long as the state variables x_i satisfy the difference equation (3), irrespective of the values of the new variables $p_1, \ldots, p_N$. Using a chain rule of differentiation (once again), prove that the gradient of $\widehat{\mathcal{J}}$ is given by

$$\boldsymbol{\nabla}_{\mathbf{u}}\widehat{\mathcal{J}} = \sum_{i=0}^{N-1}\left[\frac{\partial g_i}{\partial \mathbf{u}} + p_{i+1}\frac{\partial f_i}{\partial \mathbf{u}}\right] + \sum_{i=1}^{N-1}\left[\frac{\partial g_i}{\partial x_i} + p_{i+1}\frac{\partial f_i}{\partial x_i} - p_i\right]\frac{\partial x_i}{\partial \mathbf{u}} + \left[\frac{\partial g_N}{\partial x_N} - p_N\right]\frac{\partial x_N}{\partial \mathbf{u}}.$$

Conclude that, for some suitable choice of the *adjoint variables* p_i (the so-called *adjoint equations*), the gradient of the objective function $\mathcal{J}$ can be calculated as

$$\nabla_{u_k}\mathcal{J} = \frac{\partial g_k}{\partial u_k} + p_{k+1}\frac{\partial f_k}{\partial u_k}, \tag{6}$$

for each $k = 1, \ldots, N-1$.

Solution. Applying a chain rule of differentiation to the augmented objective function gives

$$\begin{aligned}
\boldsymbol{\nabla}_{\mathbf{u}} \widehat{\mathcal{J}} &= \frac{\partial g_N}{\partial x_N}\frac{\partial x_N}{\partial \mathbf{u}} + \sum_{i=0}^{N-1}\left(\frac{\partial g_i}{\partial x_i}\frac{\partial x_i}{\partial \mathbf{u}} + \frac{\partial g_i}{\partial \mathbf{u}} + p_{i+1}\left[\frac{\partial f_i}{\partial x_i}\frac{\partial x_i}{\partial \mathbf{u}} + \frac{\partial f_i}{\partial \mathbf{u}} - \frac{\partial x_{i+1}}{\partial \mathbf{u}}\right]\right) \\
&= \frac{\partial g_N}{\partial x_N}\frac{\partial x_N}{\partial \mathbf{u}} + \sum_{i=0}^{N-1}\left(\frac{\partial g_i}{\partial x_i}\frac{\partial x_i}{\partial \mathbf{u}} + \frac{\partial g_i}{\partial \mathbf{u}} + p_{i+1}\left[\frac{\partial f_i}{\partial x_i}\frac{\partial x_i}{\partial \mathbf{u}} + \frac{\partial f_i}{\partial \mathbf{u}}\right]\right) - \sum_{i=1}^{N} p_i \frac{\partial x_i}{\partial \mathbf{u}} \\
&= \left(\frac{\partial g_N}{\partial x_N} - p_N\right)\frac{\partial x_N}{\partial \mathbf{u}} + \sum_{i=0}^{N-1}\left(\frac{\partial g_i}{\partial \mathbf{u}} + p_{i+1}\frac{\partial f_i}{\partial \mathbf{u}}\right) + \sum_{i=1}^{N-1}\left(\frac{\partial g_i}{\partial x_i} + p_{i+1}\frac{\partial f_i}{\partial x_i} - p_i\right)\frac{\partial x_i}{\partial \mathbf{u}} \\
&\quad + \left(\frac{\partial g_0}{\partial x_0} + p_1\frac{\partial f_0}{\partial x_0}\right)\frac{\partial x_0}{\partial \mathbf{u}}.
\end{aligned}$$

Observe that we are free to choose the adjoint variables $p_1, \ldots, p_N$ as we wish, i.e., we can select $p_1, \ldots, p_N$ so as to cancel as many terms as possible in the previous equation:

$$p_N = \frac{\partial g_N}{\partial x_N} \tag{7}$$

$$p_i = p_{i+1}\frac{\partial f_i}{\partial x_i} + \frac{\partial g_i}{\partial x_i} \qquad i = 1, \ldots, N-1. \tag{8}$$

Noting also that $\frac{\partial x_0}{\partial \mathbf{u}} = \mathbf{0}$ (see question I-b above), we get

$$\boldsymbol{\nabla}_{\mathbf{u}} \widehat{\mathcal{J}} = \sum_{i=0}^{N-1}\left(\frac{\partial g_i}{\partial \mathbf{u}} + p_{i+1}\frac{\partial f_i}{\partial \mathbf{u}}\right),$$

Since $\frac{\partial f_i}{\partial u_k} = \frac{\partial g_i}{\partial u_k} = 0$ for $i \neq k$, we finally obtain

$$\nabla_{u_k} \widehat{\mathcal{J}} = \nabla_{u_k} \mathcal{J} = \frac{\partial g_k}{\partial u_k} + p_{k+1}\frac{\partial f_k}{\partial u_k}, \tag{9}$$

for each $k = 0, \ldots, N-1$.

(b) Evaluating the gradient $\boldsymbol{\nabla}_{\mathbf{u}}\mathcal{J}$ from Eq. (6) is often referred to as the *adjoint* or *reverse sensitivity approach* in the literature. Propose an algorithm that calculates $\nabla_{\mathbf{u}}\mathcal{J}$ based on this approach. How many finite difference equations is it necessary to solve for evaluating the gradient of the objective function? And how can we expect the (computational) cost of a gradient evaluation to grow with the number N of control variables? Conclude.

Solution. An estimate of the gradient $\nabla_{\mathbf{u}}\mathcal{J}$ is obtained as

$$\nabla_{\mathbf{u}}\mathcal{J} = \begin{pmatrix} \frac{\partial g_0}{\partial u_0} + p_1\frac{\partial f_0}{\partial u_0} \\ \vdots \\ \frac{\partial g_{N-1}}{\partial u_{N-1}} + p_N\frac{\partial f_{N-1}}{\partial u_{N-1}} \end{pmatrix},$$

and requires that (i) the state variables x_i, $i = 0, \ldots, N$, be calculated forward in time by integrating the difference equation (3) from the initial condition (4), and (ii) the adjoint variables p_i, $i = 1, \ldots, N$, be calculated backward in time by integrating the difference equation (8) from the terminal condition (7). Thus, an algorithm calculating $\nabla_{\mathbf{u}}\mathcal{J}$ based on the adjoint approach is as follows:

State Integration Step

Integrate the difference equation (3) forward in time, from the initial condition (4); store the resulting states $x_0, \ldots, x_N$.

Adjoint Integration Step

Integrate the difference equation (8) backward in time, from the terminal condition (7), by using the state values $x_0, \ldots, x_N$ stored during the forward pass; store the resulting adjoints $p_1, \ldots, p_N$.

Gradient Calculation Step

Calculate the components $\nabla_{u_1}\mathcal{J}, \ldots, \nabla_{u_{N-1}}\mathcal{J}$ as given in (9), based on the stored states $x_0, \ldots, x_N$ and adjoints $p_1, \ldots, p_N$.

Overall, each evaluation of $\nabla_{\mathbf{u}}\mathcal{J}$ requires that 2 difference equations only be integrated, one forward in time, the other backward in time. That is, the increase in the computational cost of a gradient evaluation is *independent* of the number N of control variables. It should be clear that the adjoint approach is better suited than the sensitivity approach for gradient calculation in those problems having a great number of control variables. The inconvenient in comparison to the forward mode of sensitivity analysis (see question I-c) is that *all* the state values $x_0, \ldots, x_N$ *must* be stored during the forward pass, since they are used in the backward pass, and this may require a lot of storage for those problems for many states and/or time stages.

III) Application to a Reservoir Regulation Problem.

Let x_i denote the volume of water held in a reservoir at the ith of N time periods. The volume x_i evolves according to

$$x_{i+1} \; = \; x_i - u_i, \quad i = 0, \ldots, N-1; \quad x_0 = 2, \tag{10}$$

where u_i is water used for some productive purpose in period i. The volume x_i can be viewed as the state, and the outflow u_i as the control.

For this problem, we define the objective function to be minimized as

$$\mathcal{J}(\mathbf{u}) \; = \; (x_N - 1)^2 + \sum_{i=0}^{N-1} \left[(x_i - 1)^2 - u_i \right].$$

That is, we want to find the control $\mathbf{u}$ that keeps the volume in the reservoir close to 1 in each time period, and at the same time, maximize the outflow u_i of water (e.g., for electric power generation). No restriction is placed on the volume x_i in the reservoir, nor on the outflow u_i.

(a) Find a mathematical expression giving the state x_i as a function of the outflows $u_0, \ldots, u_{i-1}$ in previous time periods. Conclude that the objective function is convex in $\mathbb{R}^N$.

Solution. A mathematical expression giving the state x_i as a function of the outflows $u_0, \ldots, u_{i-1}$, at each $i = 1, \ldots, N$, is easily obtained as

$$x_i \; = \; 2 - \sum_{j=1}^{i-1} u_j.$$

Note that the x_i's are affine function of the u_j's. Moreover, $\mathcal{J}$ is a convex function of the x_i's and an affine function of the $u_j's$. Therefore, the substitution of the x_i's as functions of the u_j's (composition of a convex function with an affine function) in $\mathcal{J}$ yields a convex function of the u_j's. Overall, the problem is convex and unconstrained, and any stationary point of that function is therefore a global minimum of the problem.

(b) Derive the difference equation and the associated terminal condition satisfied by the adjoint variables p_k, $k = 0, \ldots, N$, and use it to obtain an expression of the gradient $\boldsymbol{\nabla}_{\mathbf{u}}\mathcal{J}$. Conclude that stationary points for this problem are those satisfying the conditions

$$p_k^\star \;=\; -1, \quad \forall k = 1, \ldots, N.$$

Solution. For this problem, the adjoint equation (8) together with its associated terminal condition (7) specializes as

$$p_N \;= 2(x_N - 1) \tag{11}$$

$$p_i \;= p_{i+1} + 2(x_i - 1), \quad i = 1, \ldots, N-1. \tag{12}$$

According to (9), the gradient of the objective function can thus be calculated as

$$\nabla_{u_k}\mathcal{J} \;=\; -1 - p_{k+1},$$

for each $k = 0, \ldots, N-1$. For this problem, stationary controls, and hence globally optimal controls (by convexity), are those satisfying the conditions

$$p_k^\star \;=\; -1, \quad \forall k = 1, \ldots, N. \tag{13}$$

(c) Show that the foregoing stationarity conditions impose

$$x_i^\star = 1, \quad i = 1, \ldots, N-1, \quad \text{and} \quad x_N^\star = \frac{1}{2}.$$

Conclude that the *unique* optimal control sequence for that problem is

$$\mathbf{u}^\star = \left(1, 0, \ldots, 0, \frac{1}{2}\right)^\mathsf{T}.$$

Solution. From (11), we get

$$x_N^\star \;=\; 1 + \frac{1}{2}p_N^\star \;=\; \frac{1}{2}.$$

Next, (12) with (13) yield

$$x_i^\star \;=\; 1, \quad \forall k = 1, \ldots, N.$$

Re-injecting the optimal state values into (10) uniquely provides the optimal controls as

$$u_0^\star = 1, \quad u_i^\star = 0,\; i = 1, \ldots, N-2, \quad \text{and} \quad u_{N-1}^\star = \frac{1}{2}.$$