ECE 403/503 Solutions to Assignment 3

Problems for this assignment: 2.1, 2.3, 2.4, and 2.5(b), (d).

2.1 (2 points) As illustrated in Figure P2.1, we are given a smooth function $h(x)$ over interval $[0,\ 1]$ where any point x in the interval is associated with a shaded region. The problem we consider here is to find the "sweet spot" $\boldsymbol{x}^*$ in interval $[0,\ 1]$ at which the area of the shaded region reaches maximum.

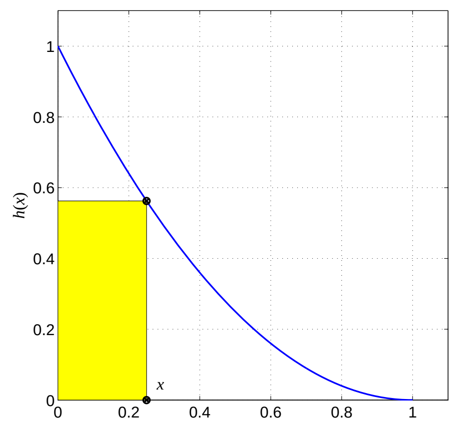

Figure P2.1 for Problem 2.1.

(a) Formulate the problem as a standard constrained minimization problem (i.e. in the form of Eq. (1.1)).

Solution

Clearly, the area of the shaded region in Fig. P2.1 is equal to $x \cdot h(x)$ where $x \in [0,1]$. Therefore, the problem at hand can be formulated as a constrained minimization problem

$$\begin{aligned} &\text{minimize} \quad -x \cdot h(x) \\ &\text{subject to:} \quad 0 \le x \le 1 \end{aligned}$$

or, in standard problem formulation as

$$\begin{aligned} &\text{minimize} \quad -x \cdot h(x) \\ &\text{subject to:} \quad c_1(x) = -x \le 0 \\ &\qquad\qquad\quad\ \ c_2(x) = x - 1 \le 0 \end{aligned}$$

■

(b) Suppose function $h(x)$ represents the straight line connecting point $[0\ \ 1]^T$ and $[1\ \ 0]^T$. Solve the problem obtained from part (a). A *numerical* solution is required.

Solution

Function $h(x)$ in this case assumes the form $h(x) = -x + 1$ and hence the objective function is given by

$$f(x) = -x \cdot h(x) = x^2 - x = (x-0.5)^2 - 0.25$$

which reaches its absolute minimum at $x^* = 0.5$. Since x^* also satisfies the constraints, it is the solution point. ■

(c) Solve the problem obtained from part (a) where $h(x) = x^2 - 2x + 1$. A *numerical* solution is required.

Solution

The objective function in this case is given by

$$f(x) = -x \cdot h(x) = -x^3 + 2x^2 - x$$

whose 1st-order derivative is equal to $f'(x) = -3x^2 + 4x - 1$. By setting $f'(x) = 0$ and solving the equation, we obtain two candidate points: $x_1 = \frac{1}{3}$ and $x_2 = 1$. The 2nd-order derivative of $f(x)$ is found to be $f''(x) = -6x + 4$, hence $f''(x_1) > 0$ and $f''(x_2) < 0$, so x_1 is the minimizer of the corresponding unstrained problem. Since $x_1 = \frac{1}{3}$ also satisfies the constraints, we conclude that

$x_1 = \frac{1}{3}$ is the solution point. ■

(d) For an arbitrary $h(x)$, derive an equation that the sweet spot $\boldsymbol{x}^*$ must satisfy.

Solution

By inspection we know that the end points of the unit interval cannot be the sweet spot because the shaded area for point $x = 0$ and $x = 1$ is zero. Therefore, the minimizer must lie strictly inside interval [0, 1]. Consequently, the 1st-order derivative of the objective function at the sweet spot must be zero, namely,

$$h(x) + x \cdot h'(x) = 0$$

■

2.3 (3 points) Compute gradient and Hessian of the functions given below.

(a) $f(x_1, x_2) = x_1^2 - 2x_2 \sin x_1 + 100$

Solution

The gradient is given by

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 2x_2 \cos x_1 \\ -2\sin x_1 \end{bmatrix}$$

and Hessian is given by

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 + 2x_2 \sin x_1 & -2\cos x_1 \\ -2\cos x_1 & 0 \end{bmatrix}$$

(b) $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{b} + \kappa$ where $\boldsymbol{H} \in R^{n\times n}$, $\boldsymbol{b} \in R^{n\times 1}$, and $\kappa \in R^{1\times 1}$ are known data.

Solution

Let us denote matrix $\boldsymbol{H}$ in terms of columns and rows as

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots & \boldsymbol{h}_n \end{bmatrix} \text{ and } \boldsymbol{H} = \begin{bmatrix} \hat{\boldsymbol{h}}_1^T \\ \hat{\boldsymbol{h}}_2^T \\ \vdots \\ \hat{\boldsymbol{h}}_n^T \end{bmatrix}$$

and denote $\boldsymbol{b}$ as

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

By definition, the partial derivative $\dfrac{\partial f(\boldsymbol{x})}{\partial x_i}$ is obtained by taking the limit

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \lim_{\delta \to 0} \frac{f(\boldsymbol{x} + \delta \boldsymbol{e}_i) - f(\boldsymbol{x})}{\delta} \quad \text{where } \boldsymbol{e}_i \text{ is the ith column of the identity matrix}$$

Now we compute

$$\begin{aligned} f(\boldsymbol{x} + \delta \boldsymbol{e}_i) &= \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H}\boldsymbol{x} + \frac{1}{2}\delta \boldsymbol{e}_i^T \boldsymbol{H}\boldsymbol{x} + \frac{1}{2}\delta \boldsymbol{x}^T \boldsymbol{H} e_i + \frac{1}{2}\delta^2 \boldsymbol{e}_i^T \boldsymbol{H}\boldsymbol{e}_i + \boldsymbol{x}^T \boldsymbol{b} + \delta \boldsymbol{e}_i^T \boldsymbol{b} + \kappa \\ &= \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H}\boldsymbol{x} + \frac{1}{2}\delta \hat{\boldsymbol{h}}_i^T \boldsymbol{x} + \frac{1}{2}\delta \boldsymbol{h}_i^T \boldsymbol{x} + \frac{1}{2}\delta^2 h_{i,i} + \boldsymbol{x}^T \boldsymbol{b} + \delta b_i + \kappa \\ &= f(\boldsymbol{x}) + \frac{1}{2}\delta\left(\hat{\boldsymbol{h}}_i^T + \boldsymbol{h}_i^T\right)\boldsymbol{x} + \frac{1}{2}\delta^2 h_{i,i} + \delta b_i \end{aligned}$$

Hence

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \lim_{\delta \to 0} \frac{f(\boldsymbol{x} + \delta \boldsymbol{e}_i) - f(\boldsymbol{x})}{\delta} = \frac{1}{2}\left(\hat{\boldsymbol{h}}_i^T + \boldsymbol{h}_i^T\right)\boldsymbol{x} + b_i$$

which leads to

$$\nabla f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{H} + \boldsymbol{H}^T)\boldsymbol{x} + \boldsymbol{b}$$

If $\boldsymbol{H}$ is symmetric, then

$$\nabla f(\boldsymbol{x}) = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{b}$$

Using the result obtained above, the Hessian can be computed by definition as

$$\nabla^2 f(\boldsymbol{x}) = \nabla\left(\nabla f(\boldsymbol{x})^T\right) = \nabla\left(\frac{1}{2}\boldsymbol{x}^T(\boldsymbol{H} + \boldsymbol{H}^T)\right) = \frac{1}{2}(\boldsymbol{H} + \boldsymbol{H}^T)$$

If $\boldsymbol{H}$ is symmetric, then we have

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{H}$$

■

(c) $f(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \ln\left(1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}_i}\right)$ where variable $\boldsymbol{\theta} \in R^{n\times 1}$, $\ln(\cdot)$ denotes natural logarithm, and $\{\boldsymbol{x}_i$ for $i = 1, 2, \ldots, N\}$ with $\boldsymbol{x}_i \in R^{n\times 1}$ is a known data set.

Solution

The gradient of $f(\boldsymbol{\theta})$ is given by

$$\nabla f(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N}\frac{e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i}}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i}}\boldsymbol{x}_i$$

from which the Hessian can be evaluated as

$$\nabla^2 f(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\frac{e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i}}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i}}\boldsymbol{x}_i\boldsymbol{x}_i^T - \frac{1}{N}\sum_{i=1}^{N}\frac{e^{-2\boldsymbol{\theta}^T\boldsymbol{x}_i}}{(1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i})^2}\boldsymbol{x}_i\boldsymbol{x}_i^T = \frac{1}{N}\sum_{i=1}^{N}\frac{e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i}}{(1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}_i})^2}\boldsymbol{x}_i\boldsymbol{x}_i^T$$

■

2.4 (3 points) Point $\boldsymbol{x}$ is called a *stationary point* of function $f(\boldsymbol{x})$ if $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$. Find and classify the stationary points of the following functions as minimizer, maximizer, or none of the above (in that case the stationary point is called a *saddle point*):

(a) $f(\boldsymbol{x}) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$

Solution

We write the objective function in question as

$$\begin{aligned} f(\boldsymbol{x}) &= 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4 \\ &= x_1^2(x_1^2 + 2x_1 + 1) + (x_1^2 - 2x_1x_2 + x_2^2) \\ &= x_1^2(x_1+1)^2 + (x_1 - x_2)^2 \end{aligned}$$

from which we see it is always nonnegative. By setting its gradient to zero, i.e.,

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 4x_1^3 + 6x_1^2 + 4x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

we obtain

$$\begin{aligned} &x_2 = x_1 \\ &4x_1^3 + 6x_1^2 + 2x_1 - 2x_2 = 0 \end{aligned}$$

that leads to an equation we shall solve:

$$2x_1^3 + 3x_1^2 + x_1 = 0$$

The solutions of the above equation are 0, –0.5, and –1, this in combination with $x_2 = x_1$ gives three stationary points of the objective function as

$$\boldsymbol{x}_a = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \boldsymbol{x}_b = \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix},\ \boldsymbol{x}_c = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

To proceed, we compute the Hessian of the objective function

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 + 12x_1 + 4 & -2 \\ -2 & 2 \end{bmatrix}$$

At $\boldsymbol{x}_a$,

$$\nabla^2 f(\boldsymbol{x}_a) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$$

which is positive definite because both eigenvalues, 5.2361 and 0.7639, are positive. Hence $\boldsymbol{x}_a$ safisfies the 2nd–order sufficient conditions and it is a minimizer.

At $\boldsymbol{x}_b$,

$$\nabla^2 f(\boldsymbol{x}_b) = \begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix}$$

which is indefinite because its eigenvalues, 3.5616 and –0.5616, have mixed signs. Hence $\boldsymbol{x}_b$ is neither a minimizer nor a maximize (called a saddle point).

At $\boldsymbol{x}_c$,

$$\nabla^2 f(\boldsymbol{x}_c) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$$

which is identical to $\nabla^2 f(\boldsymbol{x}_a)$ and is known to be positive definite. Hence $\boldsymbol{x}_c$ satisfies the 2nd–order sufficient conditions and it is a minimizer. ■

(b) $f(\boldsymbol{x}) = x_1^2 x_2^2 - 4x_1^2 x_2 + 4x_1^2 + 2x_1 x_2^2 + x_2^2 - 8x_1 x_2 + 8x_1 - 4x_2$

Solution

We express the objective function in a more compact form as

$$\begin{aligned} f(\boldsymbol{x}) &= x_1^2 x_2^2 - 4x_1^2 x_2 + 4x_1^2 + 2x_1 x_2^2 + x_2^2 - 8x_1 x_2 + 8x_1 - 4x_2 \\ &= x_1^2 (x_2 - 2)^2 + 2x_1 (x_2 - 2)^2 + (x_2 - 2)^2 - 4 \\ &= (x_1 + 1)^2 (x_2 - 2)^2 - 4 \end{aligned}$$

It immediately follows that the minimum value of the objective function is –4.

The gradient of the objective function is given by

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 + 1)(x_2 - 2)^2 \\ 2(x_1 + 1)^2 (x_2 - 2) \end{bmatrix}$$

By setting the gradient to zero, the stationary points of $f(\boldsymbol{x})$ are found to be

$$\boldsymbol{x}^* = \begin{bmatrix} x_1 \\ 2 \end{bmatrix} \quad \text{with } x_1 \text{ arbitrary}$$

and

$$\boldsymbol{x}^{**} = \begin{bmatrix} -1 \\ x_2 \end{bmatrix} \quad \text{with } x_2 \text{ arbitrary.}$$

Clearly, at any stationary point characterized above the objective function assumes its minimum value –4, we conclude that all stationary points are global minimizers. ■

(c) $f(\boldsymbol{x}) = (x_1^2 - x_2)^2 + x_1^5$

Solution

We compute its gradient and Hessian

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 4x_1(x_1^2 - x_2) + 5x_1^4 \\ -2(x_1^2 - x_2) \end{bmatrix}, \quad \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 - 4x_2 + 20x_1^3 & -4x_1 \\ -4x_1 & 2 \end{bmatrix}$$

By setting $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$, we obtain $\boldsymbol{x}^* = \boldsymbol{0}$ as the only stationary point for the objective function. At $\boldsymbol{x}^*$, the Hessian becomes

$$\nabla^2 f(\boldsymbol{x}^*) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

which is positive semidefinite. This means we need a further analysis before making conclusions.

Let $\boldsymbol{\delta} = \begin{bmatrix} \delta_1 & \delta_2 \end{bmatrix}^T$ with δ_1 and δ_2 small in magnitude but otherwise arbitrary. We compute

$$f(\boldsymbol{x}^* + \boldsymbol{\delta}) - f(\boldsymbol{x}^*) = f(\boldsymbol{\delta}) = (\delta_1^2 - \delta_2) + \delta_1^5$$

hence

$$f(\boldsymbol{x}^* + \boldsymbol{\delta}) - f(\boldsymbol{x}^*) > 0 \ \text{ when } \ \delta_2 = \delta_1^2 \ \text{ and } \ \delta_1 > 0$$
$$f(\boldsymbol{x}^* + \boldsymbol{\delta}) - f(\boldsymbol{x}^*) < 0 \ \text{ when } \ \delta_2 = \delta_1^2 \ \text{ and } \ \delta_1 < 0$$

i.e.,

$$f(\boldsymbol{x}^* + \boldsymbol{\delta}) > f(\boldsymbol{x}^*) \ \text{ when } \ \delta_2 = \delta_1^2 \ \text{ and } \ \delta_1 > 0$$
$$f(\boldsymbol{x}^* + \boldsymbol{\delta}) < f(\boldsymbol{x}^*) \ \text{ when } \ \delta_2 = \delta_1^2 \ \text{ and } \ \delta_1 < 0$$

Based on above analysis, we conclude that $\boldsymbol{x}^*$ is neither a minimizer nor a maximizer and hence is a saddle point. ■

2.5 (2 points) Function $f(\boldsymbol{x})$ is convex in a region if its Hessian is positive semidefinite in that region. Function $f(\boldsymbol{x})$ is concave if $-f(\boldsymbol{x})$ is convex. Determine whether the following functions are convex or concave or else:

(b) $f(\boldsymbol{x}) = x_1^2 + 2x_2^2 + 2x_3^2 + x_4^2 - x_1x_2 + x_1x_3 - 2x_2x_4 + x_1x_4$

Solution

The Hessian of $f(\boldsymbol{x})$ is found to be

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & -1 & 1 & 1 \\ -1 & 4 & 0 & -2 \\ 1 & 0 & 4 & 0 \\ 1 & -2 & 0 & 2 \end{bmatrix}$$

whose eigenvalues are equal to 5.8192, 4.2427, 1.3281, and 0.6099. Since all eigenvalues of the Hessian are positive, $\nabla^2 f(\boldsymbol{x})$ is positive definite and hence $f(\boldsymbol{x})$ is a convex function. ■

(d) $f(\hat{\boldsymbol{w}}) = \frac{1}{N}\sum_{i=1}^{N} \ln\left(1 + e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i}\right)$ where variable $\hat{\boldsymbol{w}} \in R^{n \times 1}$, $\ln(\cdot)$ denotes natural logarithm,

$\hat{\boldsymbol{x}}_i \in R^{n\times 1}$ and $y_i \in R^{1\times 1}$ are known data.

Solution

The function is question is practically the same as the one in Prob. 2.3c. By following the solution of Prob. 2.3c, the Hessian of $f(\hat{\boldsymbol{w}})$ is is found to be

$$\nabla^2 f(\hat{\boldsymbol{w}}) = \frac{1}{N}\sum_{i=1}^{N} \frac{y_i^2 e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i}}{(1+e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i})^2} \hat{\boldsymbol{x}}_i \hat{\boldsymbol{x}}_i^T$$

We show $\nabla^2 f(\hat{\boldsymbol{w}})$ is positive semidefinite by proving that $\boldsymbol{v}^T \nabla^2 f(\hat{\boldsymbol{w}})\boldsymbol{v} \ge 0$ for any $\boldsymbol{v}$:

$$\begin{aligned}
\boldsymbol{v}^T \nabla^2 f(\hat{\boldsymbol{w}})\boldsymbol{v} &= \boldsymbol{v}^T \left(\frac{1}{N}\sum_{i=1}^{N} \frac{y_i^2 e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i}}{(1+e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i})^2} \hat{\boldsymbol{x}}_i \hat{\boldsymbol{x}}_i^T \right) \boldsymbol{v} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{y_i^2 e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i}}{(1+e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i})^2} \boldsymbol{v}^T \hat{\boldsymbol{x}}_i \hat{\boldsymbol{x}}_i^T \boldsymbol{v} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{y_i^2 e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i}}{(1+e^{-y_i \hat{\boldsymbol{w}}^T \hat{\boldsymbol{x}}_i})^2} \left(\boldsymbol{v}^T \hat{\boldsymbol{x}}_i\right)^2 \ge 0
\end{aligned}$$

Therefore, function $f(\hat{\boldsymbol{w}})$ is convex. ■