16.888/ESD 77 Multidisciplinary System Design Optimization: Assignment 2 Part a) Solution

a1) Design of Experiments

a1-a)

Experiment #	Mean (ft)	Variance (ft^2)
1	13.9	10.9
2	12.6	6.5
3	12.9	5.4
4	12.9	7.8
5	12.4	5.3
6	**17.7**	**26.8**
7	12.1	6.1
8	13.3	18.7
9	15.1	24.5

Overall mean range, R = 13.6 ft. The variance is calculated using the unbiased estimate (i.e., variance = $s_{n-1}^2 = \frac{\sum_n (J-\bar{J})^2}{n-1}$ where 'n' is the number of experiments).

a1-b)

The design variable settings and their main effects table is shown below:

Setting	Effects
A1	-0.5
A2	**0.7**
A3	-0.1
B1	-0.7
B2	-0.9
B3	**1.6**
C1	**1.3**
C2	-0.1
C3	-1.2
D1	0.2
D2	**0.5**
D3	-0.6

a1-c)

From the above table, the optimal airplane has settings (A2, B3, C1, D2) and corresponds to experiment #6. It has the highest mean range of 17.7 ft but also has a high variance of 26.8 ft^2.

a1-d)
The overall average range (average across all experiments) is J_{mean} = 13.6 ft. Now adding the effects of the variable settings for the optimal airplane, the predicted range is J = J_{mean} + (0.7 + 1.6 + 1.3 +0.5) ft = 17.7 ft. This corresponds to the mean of experiment #6.

a1-e)

Flight:	**1**	**2**	**3**	**4**	**5**	**Mean**	**Variance**
Distance (ft):	16.5	19.75	22.5	18.6	17.9	19.05	5.105

The mean of the test flights is 19.05 ft with a variance of 5.1ft and the prediction was 17.7 feet. With a variance that large the mean of the test flight and the prediction can be considered the same or at least similar. There was considerable experimental variation during the tests as some airplanes flew straight, others flew curved paths. Therefore, with the large amount of experimental variation the prediction seems supported by experiment

a1-f)
The optimal airplane setting becomes the baseline design for conducting further parameter study. In a parameter study, only one factor is changed at a time, keeping all other variables at the baseline setting.
The number of experimental points = 1+n*(l-1) =1+4*2=9.

Experiment #	A	B	C	D
1 (baseline)	**A2**	**B3**	**C1**	**D2**
2	A1	B3	C1	D2
3	A3	B3	C1	D2
4	A2	B1	C1	D2
5	A2	B2	C1	D2
6	A2	B3	C2	D2
7	A2	B3	C3	D2
8	A2	B3	C1	D1
9	A2	B3	C1	D3

Except the base design, none of the previous designs is included in this table. Therefore each of the 8 new experiments would potentially lead to enhanced understanding of the design space, including possible interactions. This might lead to improved estimate on mean and reduce the variability.

a1-g)
A larger variance indicates a wider spread of results about a mean value and reduces the confidence one has in the mean value as a predictor. This

essentially means that the mean is not a good enough predicted of the expected performance.

Standard deviation, which is the square root of the variance, can be used to define confidence bounds on the expected range and would serve as an indicator while selecting an optimal design setting for further exploration.

a2) Gradient Based Optimization

(a):

The objective function is:

$$f(x_1, x_2) = x_1^4 - x_1^2 x_2 + x_2^2 + \frac{1}{2} x_1^2 \quad (1)$$

The corresponding gradient vector and hessian matrix in symbolic for are;

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1^3 - 2x_1 x_2 + x_1 \\ -x_1^2 + 2x_2 \end{bmatrix} \quad (2)$$

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 12x_1^2 - 2x_2 + 1 & -2x_1 \\ -2x_1 & 2 \end{bmatrix} \quad (3)$$

(i) **Steepest Descent Method**

Starting point is, $X^0 = [2 \quad 2]^T$. The next point in the function minimizing sequence $\{X^k\}$, generated by the steepest descent method is given as:

$$X^1 = X^0 - \alpha \nabla f(X^0) \quad (4)$$

where α = step length and $S^0 = -\nabla f(X^0)$ is the search direction. Using X^0 in eq. (2) and then substituting in eq. (4), we get the next point in terms of step length as $X^1 = [(2 - 26\alpha) \quad 2]^T$. We have to substitute X^1 in eq. (1):

$$f_\alpha(X^1) = (2 - 26\alpha)^4 - 2(2 - 26\alpha)^2 + 4 + \frac{1}{2}(2 - 26\alpha)^2 \quad (5)$$

This has now reduced to a single variable optimization problem in α. Equating the first derivative to zero, the stationary points are:

$$\alpha_1 = \frac{1}{13}(1 + \frac{\sqrt{3}}{4}); \; \alpha_2 = \frac{1}{13}; \; \alpha_3 = \frac{1}{13}(1 - \frac{\sqrt{3}}{4})$$

The second derivative is positive for $\alpha = \frac{1}{13}(1+\frac{\sqrt{3}}{4})$ and there the function in eq. (5) has a minimum at $\alpha = \frac{1}{13}(1+\frac{\sqrt{3}}{4}) = 0.11$.

Using this α we get $X^1 = \begin{bmatrix}-0.866 & 2\end{bmatrix}^T$.

** Check: $f(X^0) = 14$; $f(X^1) = 3.4375 < f(X^0)$. Hence the objective function value has reduced by almost three times as we moved from point X^0 to X^1.

(ii) **Newton's Method**

In the pure Newton method, the next iterate is given by;

$$X^1 = X^0 - H^{-1}(X^0)\nabla f(X^0) \quad (6)$$

The gradient vector and hessian matrix are as follows:

$$\nabla f(X^0) = \begin{bmatrix} 26 \\ 0 \end{bmatrix}$$
$$H(X^0) = \begin{bmatrix} 45 & -4 \\ -4 & 2 \end{bmatrix}$$

Let $Z = X^1 - X^0$ and we can rewrite eq. (6) as a system of linear equations:

$$H(X^0)Z = -\nabla f(X^0) \quad (7)$$

Solving eq. (7) , we get the new iterate: $X^1 = \begin{bmatrix} 48/37 \\ 22/37 \end{bmatrix} = \begin{bmatrix} 1.2973 \\ 0.5946 \end{bmatrix}$.

** Check: $f(X^1) = 3.0274 < f(X^0)$. Son the objective function value has reduced as we moved from initial guess to the first iterate.

a2-b):

$$\begin{aligned} &\text{Min.} \quad f(x_1, x_2) = x_1^2 + x_2^2 \\ &\text{s.t} \quad 2 \le x_1 \;\Rightarrow\; \underbrace{(2 - x_1)}_{g(x_1,x_2)} \le 0 \end{aligned}$$

(i) Lagrangian: $L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2) = x_1^2 + x_2^2 + \lambda(2 - x_1)$, assuming active constraint. Now notice that size of the system to be solved has increased to three, including the lagrange multiplier λ. The KKT condition for the Lagrangian is:

$$\nabla L = \begin{bmatrix} \frac{\partial L}{\partial x_1} \\ \frac{\partial L}{\partial x_2} \\ \frac{\partial L}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 \\ 2 - x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (1)$$

The solution to (1) is:

$$\begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}, \quad \lambda = 4 > 0$$

and the constraint is active. So it satisfies the complimentary slackness criteria.
The function value at the KKT point is $f^* = 4$.
In order to check if KKT point leads to a minimum, we need to look at the hessian matrix of the Lagrangian, given that the solution is feasible.
The Hessian of the Lagrangian in this case is:

$$\nabla_{xx} L = \begin{bmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \succ 0 \text{ positive definite}$$

since both eigenvalues (2,2) positive (2,2). Hence the function is minimized at the KKT point.

(ii) Use of logarithmic barrier function

The penalized objective function is:

$$\Phi_L = f(x_1, x_2) - r_p[\ln(x_1 - 2)] \quad (1)$$

where r_p is the penalty or barrier parameter. Notice that the term associated to the constraint is undefined for all infeasible points (e.g., where g(x)>0).
This represents an 'interior point' method where iteration starts inside the feasible region and tries to reach the constraint boundary from inside the feasible region.
Apply KKT condition to Φ_L in eq. (1):

$$\begin{bmatrix} \dfrac{\partial \Phi_L}{\partial x_1} \\ \dfrac{\partial \Phi_L}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$=> \begin{bmatrix} 2x_1 - \dfrac{r_p}{x_1 - 2} \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Now Φ_L is defined only for the feasible region, $x_1 \geq 2$. Hence the stationary point is given as:

$$X^*(r_p) = \begin{bmatrix} 1 + \sqrt{1 + \dfrac{r_p}{2}} \\ 0 \end{bmatrix} \quad (2)$$

The optimal point is $X^* = \lim_{r_p \to 0} X^*(r_p) = [2 \quad 0]^T$. The penalized objective function reaches the boundary from below, $\Phi_L \to f(X)$ as $r_p \to 0$.

(iii) Newton step on Lagrangian:

Assuming active constraint, the Lagrangian is:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2) = x_1^2 + x_2^2 + \lambda(2 - x_1) \quad (1)$$

The corresponding gradient and hessian are

$$\nabla L = \begin{bmatrix} \dfrac{\partial L}{\partial x_1} \\ \dfrac{\partial L}{\partial x_2} \\ \dfrac{\partial L}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 \\ 2 - x_1 \end{bmatrix}$$

$$H_L = \begin{bmatrix} \dfrac{\partial^2 L}{\partial x_1^2} & \dfrac{\partial^2 L}{\partial x_1 \partial x_2} & \dfrac{\partial^2 L}{\partial x_1 \partial \lambda} \\ \dfrac{\partial^2 L}{\partial x_2 \partial x_1} & \dfrac{\partial^2 L}{\partial x_2^2} & \dfrac{\partial^2 L}{\partial x_2 \partial \lambda} \\ \dfrac{\partial^2 L}{\partial \lambda \partial x_1} & \dfrac{\partial^2 L}{\partial \lambda \partial x_2} & \dfrac{\partial^2 L}{\partial \lambda^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

Let us assume an initial point: $X^0 = [3 \quad 1 \quad 0]^T$. Hence $\nabla L^0 = [6 \quad 2 \quad -1]^T$.

Apply Newton's method as before yields:

$$X^1 = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}$$

Comparing with part (i), one can observe that Newton's method has converged in one step. This is because the objective function here is quadratic with a constraint is linear.

a2-c):

$$\begin{aligned} &\text{Min.} \quad f(x_1, x_2) = 4x_1^2 + 12x_2^2 \\ &\text{s.t} \quad h(x_1, x_2) = x_2^2 - (x_1 - 1)^3 = 0 \end{aligned}$$

(i) The Lagrangian is:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2) = 4x_1^2 + 12x_2^2 + \lambda[x_2^2 - (x_1 - 1)^3] \quad (1)$$

The KKT condition yields the following system of equations:

$$\nabla L = \begin{bmatrix} \dfrac{\partial L}{\partial x_1} \\ \dfrac{\partial L}{\partial x_2} \\ \dfrac{\partial L}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} 8x_1 - 3\lambda(x_1 - 1)^2 \\ 24x_2 + 2\lambda x_2 \\ x_2^2 - (x_1 - 1)^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (2)$$

The system of equation (2) does not have a consistent real solution.

This situation is associated to *linear independence constraint qualification* (LICQ). In the KKT condition, there is an inherent assumption that the constraint jacobian (in case of multiple constraints) be of full-rank or the constraint vector (in case of a single constraint) is non-zero at the KKT point. [Recall the KKT condition, $\underbrace{\nabla f(X)}_{nx1} + \underbrace{\nabla h(X)^T}_{nxm} \underbrace{\lambda}_{mx1} = \underbrace{0}_{nx1} \Rightarrow \lambda = -[\nabla h(X)^T]^{-1} \nabla f(X)$ and inverse of the constraint jacobian should exist].

Here, the actual optima lies at (1,0) but $\nabla h = \begin{bmatrix} -3(x_1 - 1)^2 & 2x_2 \end{bmatrix}^T$ reduces to a zero vector at (1,0) and the assumption of KKT condition is violated in this case.

$$\begin{bmatrix} \dfrac{\partial f(1,0)}{\partial x_1} \\ \dfrac{\partial f(1,0)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \dfrac{\partial h(1,0)}{\partial x_1} \\ \dfrac{\partial h(1,0)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

KKT cannot be satisfied: $\begin{bmatrix} 8 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

(ii) The quadratic penalty function for this problem is:

$$\Phi_Q = \underbrace{4x_1^2 + 12x_2^2}_{f} + \rho[\underbrace{x_2^2 - (x_1 - 1)^3}_{h}]^2$$

$$=> \ \nabla \Phi_Q = \nabla f + 2\rho h \nabla h = 0 \ \text{ for stationary points} \qquad (1)$$

From KKT condition,

$$\frac{\partial \Phi_Q}{\partial x_2} = 0 => \ x_2\{\rho[x_2^2 - (x_1 - 1)^3] + 6\} = 0$$

$$=> \ x_2 = 0 \text{ or, } x_2^2 - (x_1 - 1)^3 = -\frac{6}{\rho}$$

As $\rho \to \infty$, we have $x_1 = 1 \text{ and } x_2 = 0$ as the solution.

Therefore, $X^* = \lim_{\rho \to \infty} X^*(\rho) = [1 \quad 0]^T$.

Alternative, for KKT condition $\nabla f + 2\rho h \nabla h = 0$ to be satisfied as $\rho \to \infty$, we must have $h \nabla h$ tending to zero. At (1,0), both $h \text{ and } \nabla h$ are zero (i.e., the function has a singular point).

(iii)

The contour plot of the objective function and the constraint shows that the gradient vector of the constraint at (1,0) is zero as does the function value. Since the LICQ does not hold at this point, the KKT condition is not valid at this point.

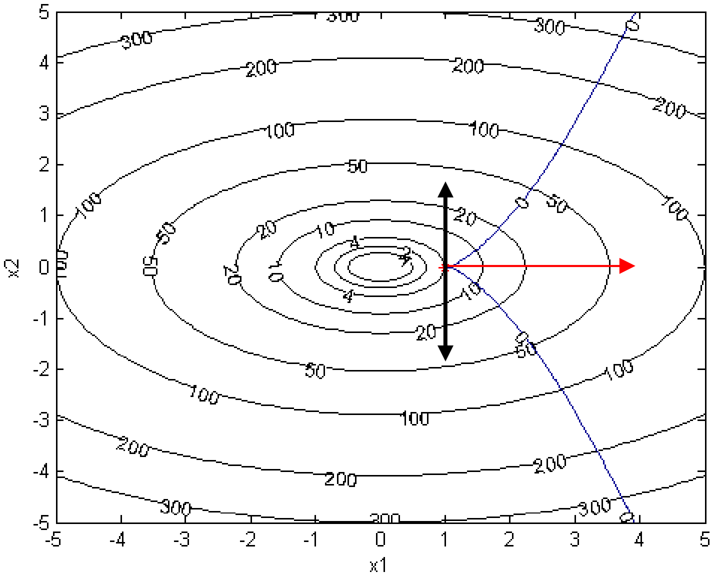

Fig. 1: Contour plots of the object function and constraint.

One can also split the equality constraint into two:

$$h_1(X) \equiv x_2 - (x_1 - 1)^{3/2} = 0$$
$$h_2(X) \equiv x_2 + (x_1 - 1)^{3/2} = 0$$

The associated constraint jacobian is:

$$\nabla h^T = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_2}{\partial x_1} \\ \frac{\partial h_1}{\partial x_2} & \frac{\partial h_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$$

This matrix is singular and its columns are linearly dependent. This violates the regularity or LICQ criteria and use of KKT conditions is invalidated.

ESD.77 / 16.888 Multidisciplinary System Design Optimization
Spring 2010