KKT Conditions: Support Vector Machines

Y Aditya, G V S S Praneeth Varma and G V V Sharma*

Contents

1 Lagrangian 1

2 Karush Kuhn-Tucker Conditions 2

***Abstract*—This manual provides a simple introduction to various concepts in optimization.**

1 Lagrangian

1.1 Plot the circles

$$f(\mathbf{x}) = (x_1 - 8)^2 + (x_2 - 6)^2 = r^2 \tag{1.1}$$

$\mathbf{x} = (x_1, x_2)^T$, for different values of r along with the line

$$g(\mathbf{x}) = x_1 + x_2 - 9 = 0 \tag{1.2}$$

using the following program. From the graph, find

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \tag{1.3}$$

$$g(\mathbf{x}) = x_1 + x_2 - 9 = 0 \tag{1.4}$$

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.1.py
```

1.2 Obtain a theoretical solution for problem 1.1 using coordinate geometry.
Solution: From (1.2) and (1.1),

$$r^2 = (x_1 - 8)^2 + (3 - x_1)^2 \tag{1.5}$$

$$= 2x_1^2 - 22x_1 + 73 \tag{1.6}$$

$$\Rightarrow r^2 = \frac{(2x_1 - 11)^2 + 5^2}{2} \tag{1.7}$$

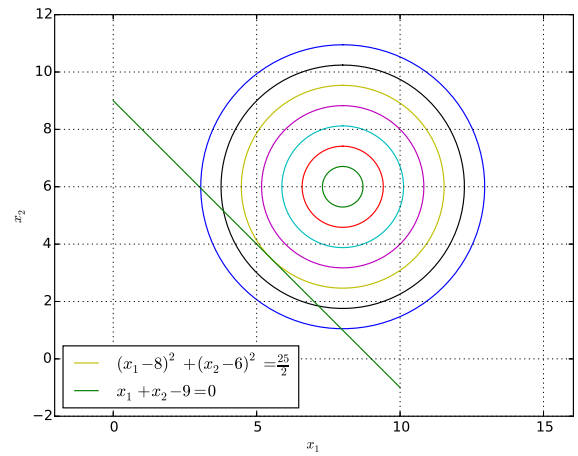

Fig. 1.1: Finding $\min_{\mathbf{x}} f(\mathbf{x})$

.

which is minium when $x_1 = \frac{11}{2}, x_2 = \frac{7}{2}$. The minimum value is $\frac{25}{2}$ and the radius $r = \frac{5}{\sqrt{2}}$.

1.3 Define

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x}) \tag{1.8}$$

and

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial \lambda} \end{pmatrix} \tag{1.9}$$

Solve the equations

$$\nabla L(\mathbf{x}, \lambda) = 0. \tag{1.10}$$

How is this related to problem 1.1? What is the sign of λ? L is known as the Lagrangian and the above technique is known as the Method of Lagrange Multipliers.

* The authors are with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

Solution: From (1.2) and (1.1),

$$L(\mathbf{x},\lambda) = (x_1-8)^2 + (x_2-6)^2 - \lambda(x_1+x_2- \tag{1.11}$$

$$\Rightarrow \nabla L(\mathbf{x},\lambda) = \begin{pmatrix} 2x_1-16-\lambda \\ 2x_2-12-\lambda \\ x_1+x_2-9 \end{pmatrix} \tag{1.12}$$

$$= \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 16 \\ 12 \\ 9 \end{pmatrix} = 0 \tag{1.13}$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} \frac{11}{2} \\ \frac{7}{2} \\ -5 \end{pmatrix} \tag{1.14}$$

using the following python script. Note that this method yields the same result as the previous exercises. Thus, λ is negative.

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.3.py
```

1.4 Modify the code in problem 1.1 to find a graphical solution for minimising

$$f(\mathbf{x}) = (x_1-8)^2 + (x_2-6)^2 \tag{1.15}$$

with constraint

$$g(\mathbf{x}) = x_1 + x_2 - 9 \geq 0 \tag{1.16}$$

Solution: This problem reduces to finding the radius of the smallest circle in the shaded area in Fig. 1.4 . It is clear that this radius is 0.

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.4.py
```

2 Karush Kuhn-Tucker Conditions

2.1 Now use the method of Lagrange multipliers to solve problem 1.4 and compare with the graphical solution. Comment.
Solution: Using the method of Lagrange multipliers, the solution is the same as the one obtained in problem 1.4, which is different from the graphical solution. This means that the Lagrange multipliers method cannot be applied blindly.

2.2 Repeat problem 2.1 by keeping $\lambda = 0$. Comment.

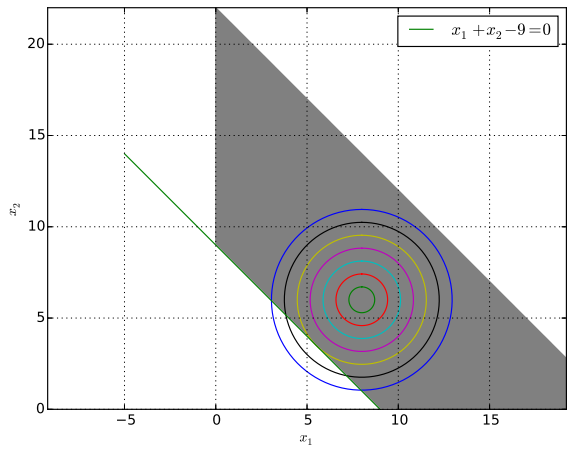

Fig. 1.4: Smallest circle in the shaded region is a point.

Solution: Keeping $\lambda = 0$ results in $x_1 = 8, x_2 = 6$, which is the correct solution. The minimum value of $f(\mathbf{x})$ without any constraints lies in the region $g(\mathbf{x}) = 0$. In this case, $\lambda = 0$.

2.3 Find a graphical solution for minimising

$$f(\mathbf{x}) = (x_1-8)^2 + (x_2-6)^2 \tag{2.1}$$

with constraint

$$g(\mathbf{x}) = x_1 + x_2 - 9 \leq 0. \tag{2.2}$$

Summarize your observations.
Solution: In Fig. 2.3, the shaded region represents the constraint. Thus, the solution is the same as the one in problem 1.4. This implies that the method of Lagrange multipliers can be used to solve the optimization problem with this inequality constraint as well. Table 2.3 summarizes the conditions for this based on the observations so far.

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.7.py
```

TABLE 2.3: Summary of conditions.

Cost	**Constraint**	λ
$f(\mathbf{x})$	$g(\mathbf{x}) = 0$	< 0
	$g(\mathbf{x}) \geq 0$	0
	$g(\mathbf{x}) \leq 0$	< 0

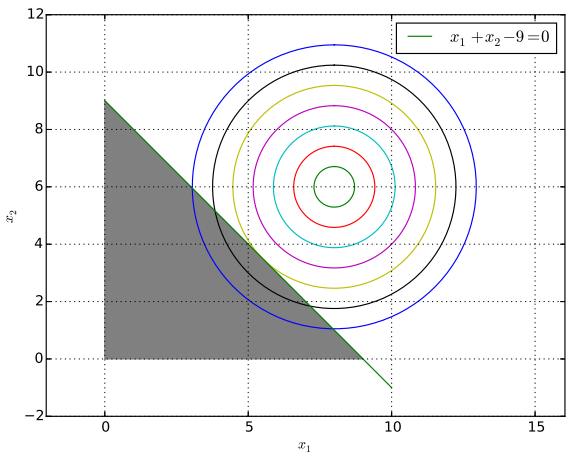

Fig. 2.3: Finding $\min_{\mathbf{x}} f(\mathbf{x})$.

2.4 Find a graphical solution for

$$\min_{\mathbf{x}} f(\mathbf{x}) = (x_1 - 8)^2 + (x_2 - 6)^2 \tag{2.3}$$

with constraint

$$g(\mathbf{x}) = x_1 + x_2 - 18 = 0 \tag{2.4}$$

Solution:

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.8.py
```

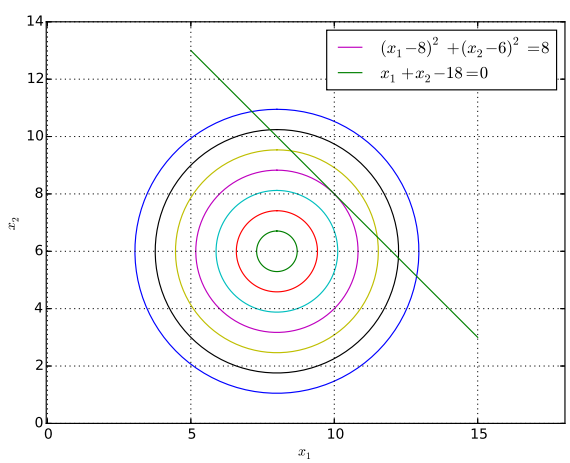

Fig. 2.4: Finding $\min_{\mathbf{x}} f(\mathbf{x})$.

2.5 Repeat problem 2.4 using the method of Lagrange mutipliers. What is the sign of λ?

Solution: From (2.3) and (2.4),

$$L(\mathbf{x}, \lambda) = (x_1 - 8)^2 + (x_2 - 6)^2 - \lambda (x_1 + x_2 - 18) \tag{2.5}$$

$$\Rightarrow \nabla L(\mathbf{x}, \lambda) = \begin{pmatrix} 2x_1 - 16 - \lambda \\ 2x_2 - 12 - \lambda \\ x_1 + x_2 - 18 \end{pmatrix} \tag{2.6}$$

$$= \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 16 \\ 12 \\ 18 \end{pmatrix} = 0 \tag{2.7}$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 10 \\ 8 \\ 4 \end{pmatrix} \tag{2.8}$$

using the following python script. Thus, λ is positive and the minimum value of f is 8.

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.9.py
```

2.6 Solve

$$\min_{\mathbf{x}} f(\mathbf{x}) = (x_1 - 8)^2 + (x_2 - 6)^2 \tag{2.9}$$

with constraint

$$g(\mathbf{x}) = x_1 + x_2 - 18 \geq 0 \tag{2.10}$$

Solution: Since the unconstrained solution is outside the region $g(\mathbf{x}) \geq 0$, the solution is the same as the one in problem 2.4.

2.7 Based on the problems so far, generalise the Lagrange multipliers method for

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad g(\mathbf{x}) \geq 0 \tag{2.11}$$

Solution: Considering $L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x})$, for $g(\mathbf{x}) = x_1 + x_2 - 18 \geq 0$ we found $\lambda > 0$ and for $g(\mathbf{x}) = x_1 + x_2 - 9 \leq 0, \lambda < 0$. A single condition can be obtained by framing the optimization problem as

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad g(\mathbf{x}) \leq 0 \tag{2.12}$$

with the Lagrangian

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x}), \tag{2.13}$$

provided

$$\nabla L(\mathbf{x}, \lambda) = 0 \Rightarrow \lambda > 0 \tag{2.14}$$

else, $\lambda = 0$.

2.8 Solve

$$\min_{\mathbf{x}} f(\mathbf{x}) = 4x_1^2 + 2x_2^2 \tag{2.15}$$

with constraints

$$g_1(\mathbf{x}) = 3x_1 + x_2 - 8 = 0 \tag{2.16}$$
$$g_2(\mathbf{x}) = 15 - 2x_1 - 4x_2 \geq 0 \tag{2.17}$$

Solution: Considering the Lagrangian

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g_1(\mathbf{x}) - \mu g_2(\mathbf{x}) \tag{2.18}$$
$$= 4x_1^2 + 2x_2^2 + \lambda(3x_1 + x_2 - 8)$$
$$- \mu(15 - 2x_1 - 4x_2), \tag{2.19}$$

$$\nabla L(\mathbf{x}, \lambda) = \begin{pmatrix} 8x_1 + 3\lambda + 2\mu \\ 4x_2 + \lambda + 4\mu \\ 3x_1 + x_2 - 8 \\ -2x_1 - 4x_2 + 15 \end{pmatrix} = 0 \tag{2.20}$$

resulting in the matrix equation

$$\Rightarrow \begin{pmatrix} 8 & 0 & 3 & 2 \\ 0 & 4 & 1 & 4 \\ 3 & 1 & 0 & 0 \\ 2 & 4 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 8 \\ 15 \end{pmatrix} \tag{2.21}$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} 1.7 \\ 2.9 \\ -3.12 \\ -2.12 \end{pmatrix} \tag{2.22}$$

using the following python script. The (incorrect) graphical solution is available in Fig. 2.8

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.12.py
```

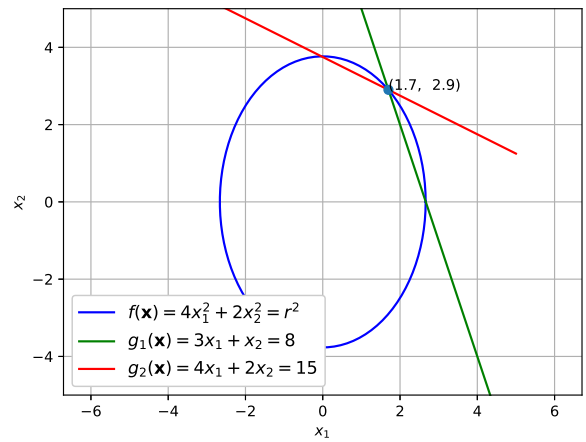

Fig. 2.8: Incorrect solution is at intersection of all curves $r = 5.33$

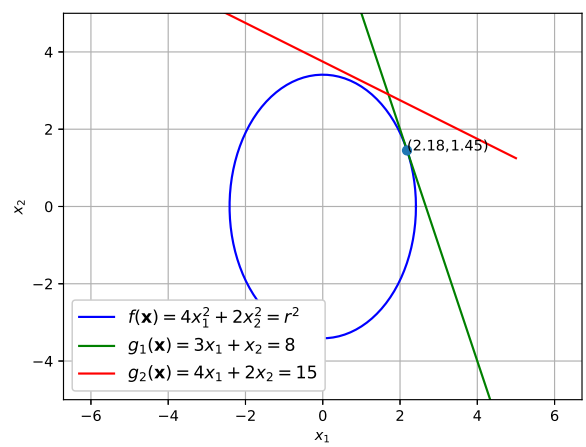

Fig. 2.9: Optimal solution is where $g_1(x)$ touches the curve $r = 4.82$

Note that $\mu < 0$, contradicting the necessary condition in (2.14).

2.9 Obtain the correct solution to the previous problem by considering $\mu = 0$.

2.10 Solve

$$\min_{\mathbf{x}} f(\mathbf{x}) = 4x_1^2 + 2x_2^2 \tag{2.23}$$

with constraints

$$g_1(\mathbf{x}) = 3x_1 + x_2 - 8 = 0 \tag{2.24}$$
$$g_2(\mathbf{x}) = 15 - 2x_1 - 4x_2 \leq 0 \tag{2.25}$$

2.11 Based on whatever you have done so far, list the steps that you would use in general for solving a convex optimization problem like (2.15) using Lagrange Multipliers. These are called Karush-Kuhn-Tucker(KKT) conditions.

Solution: For a problem defined by

$$\mathbf{x}^* = \min_{\mathbf{x}} f(\mathbf{x}) \tag{2.26}$$

Primal Feasibility :

$$\text{subject to } h_i(\mathbf{x}) = 0, \forall i = 1, .., m \tag{2.27}$$

$$\text{subject to } g_i(\mathbf{x}) \leq 0, \forall i = 1, .., n \tag{2.28}$$

the optimal solution is obtained through

$$\mathbf{x}^* = \min_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu) \tag{2.29}$$

$$= \min_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) + \sum_{i=1}^{n} \mu_i g_i(\mathbf{x}), \tag{2.30}$$

using the KKT conditions

Stationarity :

$$\Rightarrow \nabla_{\mathbf{x}} f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla_{\mathbf{x}} h_i(\mathbf{x}) + \sum_{i=1}^{n} \mu_i \nabla_{\mathbf{x}} g_i(\mathbf{x}) = 0 \tag{2.31}$$

Complementary Slackness :

$$\text{subject to } \mu_i g_i(\mathbf{x}) = 0, \forall i = 1, .., n \tag{2.32}$$

Dual Feasibility :

$$\text{and } \mu_i \geq 0, \forall i = 1, .., n \tag{2.33}$$

2.12 Maxmimize

$$f(\mathbf{x}) = \sqrt{x_1 x_2} \tag{2.34}$$

with the constraints

$$x_1^2 + x_2^2 \leq 5 \tag{2.35}$$

$$x_1 \geq 0, x_2 \geq 0 \tag{2.36}$$

2.13 Solve

$$\min_{\mathbf{x}} \quad x_1 + x_2 \tag{2.37}$$

with the constraints

$$x_1^2 - x_1 + x_2^2 \leq 0 \tag{2.38}$$

where $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Solution: Using the method of Lagrange multipliers,

$$\nabla \{f(\mathbf{x}) + \mu g(\mathbf{x})\} = 0, \quad \mu \geq 0 \tag{2.39}$$

resulting in the equations

$$2x_1\mu - \mu + 1 = 0 \tag{2.40}$$

$$2x_2\mu + 1 = 0 \tag{2.41}$$

$$x_1^2 - x_1 + x_2^2 = 0 \tag{2.42}$$

which can be simplified to obtain

$$\left(\frac{1-\mu}{2\mu}\right)^2 + \left(\frac{1}{2\mu}\right)^2 + \frac{1-\mu}{2\mu} = 0 \tag{2.43}$$

$$\Rightarrow 1 + \mu^2 - 2\mu + 1 + 2\mu(1-\mu) = 0 \tag{2.44}$$

$$\Rightarrow \mu^2 = 2, or \mu = \pm\sqrt{2} \tag{2.45}$$

From (2.15), $\mu \geq 0 \Rightarrow \mu = \sqrt{2}$. The desired solution is

$$\mathbf{x} = \begin{pmatrix} \frac{\sqrt{2}-1}{2\sqrt{2}} \\ -\frac{1}{2\sqrt{2}} \end{pmatrix} \tag{2.46}$$

Graphical solution: The constraint can be expressed as

$$x_1^2 - x_1 + x_2^2 \leq 0 \tag{2.47}$$

$$\Rightarrow \left(x_1 - \frac{1}{2}\right)^2 + x_2^2 \leq \left(\frac{1}{2}\right)^2 \tag{2.48}$$

```
wget https://raw.githubusercontent.com/
    gadepall/EE2250/master/manual/codes
    /2.15.py
```

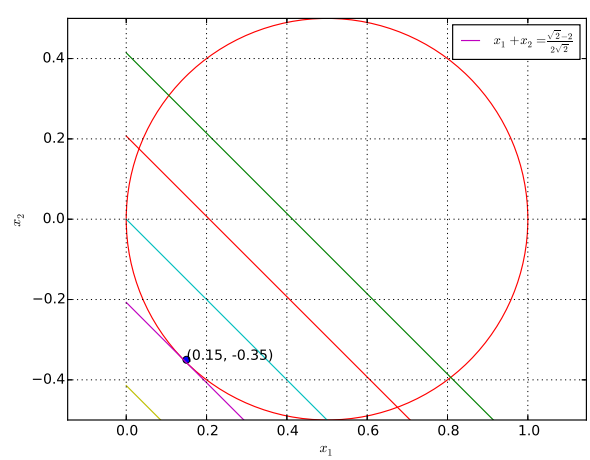

Fig. 2.13: Optimal solution is the lower tangent to the circle