
7.0 Equality Constraints: Lagrange Multipliers

Consider the minimization of a non-linear function subject to equality constraints:

$$\begin{cases} \min_{\mathbf{x}} f(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n \\ g_i(\mathbf{x}) = 0 & i = 1(1)m \end{cases} \quad (7.1)$$

where the $g_i(\mathbf{x})$ are possibly also nonlinear functions, and $m < n$ otherwise we have the possibility of an over-determined system of constraints. If the $g_i(\mathbf{x})$ functions are linear or simple, then one variable can be eliminated for each equality constraint, *i.e.* m variables can be eliminated, thus transforming the problem to an $(n-m)$ variable unconstrained minimization problem.

Consider the following example in \mathbb{R}^3 with one equality constraint:

$$\begin{cases} \min_{\mathbf{x}} f(\mathbf{x}) = x_1 x_2 x_3 \\ g(\mathbf{x}) = x_1 + x_2 + x_3 - 1 = 0 \end{cases}$$

eliminating x_3 we have

$$x_3 = 1 - x_1 - x_2$$

which when substituted back into the objective function gives us a new objective function in \mathbb{R}^2

$$\min_{\mathbf{x}} \hat{f}(\mathbf{x}) = x_1 x_2 (1 - x_1 - x_2)$$

which is now an unconstrained minimization problem.

Of course, this can't always be done easily since the equality constraints may be complicated or even implicitly defined. A general procedure for incorporating the equality constraints into the objective function was developed by *Lagrange* in 1760. In this method a new *unconstrained* problem is formed by appending the constraints to the objective function with so-called *Lagrange multipliers*. We will now describe this method.

7.1 Lagrange Multipliers

If we have an objective function in \mathbb{R}^n with m equality constraints, such as in (7.1) then we can introduce m new variables called *Lagrange multipliers*,

$$\lambda_i, i = 1(1)m \quad (7.2)$$

to create a new objective function is called the *Lagrangian*, $L(\mathbf{x}, \lambda)$, defined as

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}). \quad (7.3)$$

We now must minimize the Lagrangian over the \mathbb{R}^{n+m} space of the original variables \mathbf{x} plus the new Lagrange multipliers λ . Therefore, we have eliminated the equality constraints at the expense of increasing the dimension of our problem from \mathbb{R}^n to \mathbb{R}^{n+m} .

We can now apply the optimality conditions as before. Recall the first order *necessary condition* that the minimum be at a stationary point. Therefore, if we take the gradient of the Lagrangian function we arrive at the following necessary conditions:

$$\begin{cases} \left. \frac{\partial L}{\partial x_j} = \left[\frac{\partial f}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial g_i(\mathbf{x})}{\partial x_j} \right] \right|_{\substack{\mathbf{x} = \mathbf{x}^* \\ \lambda = \lambda^*}} = 0 & j = 1(1)n \\ \frac{\partial L}{\partial \lambda_i} = g_i(\mathbf{x}^*) = 0 & i = 1(1)m \end{cases} \quad (7.4)$$

These simultaneous equations are solved for $(\mathbf{x}^*, \lambda^*)$, that is, we have $m+n$ equations in $m+n$ unknowns. Note that the second set of these equations are just the original constraints!

Also, since at the stationary point of the Lagrangian, $(\mathbf{x}^*, \lambda^*)$, we have

$$g_i(\mathbf{x}^*) = 0$$

this means that

$$L(\mathbf{x}^*, \lambda^*) = f(\mathbf{x}^*) \quad (7.5)$$

but it is not necessarily the case that $\nabla f(\mathbf{x}^*) = \mathbf{0}$. That is, in a problem where we have equality constraints, the minimum is not necessarily found at a stationary point of the original objective

function. If $\nabla f(\mathbf{x}^*) = \mathbf{0}$ then it is because the feasible region defined by the equality constraints includes the unconstrained minimum of the function.

The use of this method can be cleared up by considering an example. Consider the following simple problem in \mathbb{R}^2 with one equality constraint:

$$\begin{cases} \min_{\mathbf{x}} f(\mathbf{x}) = \frac{1}{2}(x_1^2 + x_2^2) \\ 2x_1 - x_2 = 5 \end{cases}$$

Geometrically, the problem is to find the point of shortest distance from the origin

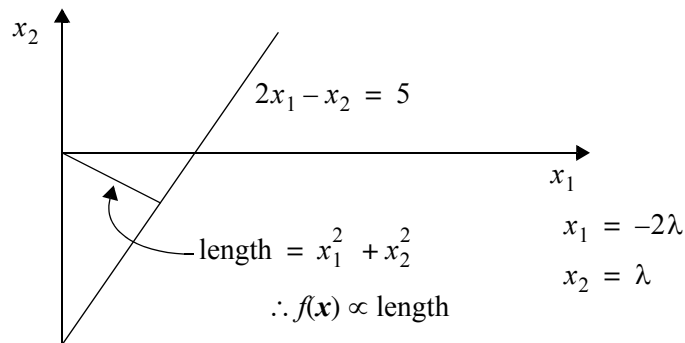


Figure 7.1 Simple equality constrained example.

Method 1: Elimination of a variable

The first method we try is to eliminate one of the variables. Therefore, solving for x_2 in terms of x_1 we have

$$x_2 = 2x_1 - 5$$

which when substituted back into the objective function, gives us a new objective function of just one variable

$$\hat{f}(x_1) = \frac{1}{2}(x_1^2 + (2x_1 - 5)^2)$$

$$\left. \frac{\partial \hat{f}}{\partial x_1} \right|_{x_1^*} = x_1^* + (2x_1^* - 5)2 = 0$$

Solving for x_1^* gives us $x_1^* = 2$, $x_2^* = -1$; that is $\mathbf{x}^* = (2, -1)$. This is a very simple example because the equality constraint was such that one of the decision variables could be easily eliminated. We will now see how the Lagrangian method can be used to solve the same problem.

Method 2: Lagrangian

We first construct the Lagrangian as

$$L(\mathbf{x}, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) + \lambda(2x_1 - x_2 - 5)$$

and then set the gradient of this function to zero:

$$\begin{cases} \frac{\partial L}{\partial x_1} = x_1^* + 2\lambda^* = 0 \\ \frac{\partial L}{\partial x_2} = x_2^* - \lambda^* = 0 \\ \frac{\partial L}{\partial \lambda} = 2x_1^* - x_2^* - 5 = 0 \end{cases}$$

This is a set of 3 equations in 3 unknowns. We first solve the first two for the Lagrange multiplier, λ^* , and then substitute into the third, giving

$$-4\lambda^* - \lambda^* - 5 = 0 \Rightarrow \lambda^* = -1$$

Once we have the Lagrange multiplier, we can easily solve for the remaining variables

$$(\mathbf{x}^*, \lambda^*) = (2, -1, -1)$$

This Lagrangian, L , is a quadratic function which can be written in matrix form as

$$L = \begin{bmatrix} 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 & x_2 & \lambda \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix}$$

from which we see that the Hessian matrix is given by

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 0 \end{bmatrix}$$

and since $|A| = -5$, A is not positive definite nor semi-definite. Therefore, the solution $(\mathbf{x}^*, \lambda^*)$ is not a minimum of $L(\mathbf{x}, \lambda)$ but \mathbf{x}^* is a minimum of the constrained function $f(\mathbf{x})$. We now

check $-A$ to see if it is positive definite. Since $|-A| = -1$, A is not a negative definite nor negative semi-definite. Therefore the solution $(\mathbf{x}^*, \lambda^*)$ is a *saddle point* of the Lagrangian function.

7.2 Quadratic Objective Functions with Linear Equality Constraints

We now consider the specialized problem wherein the objective function is a positive definite quadratic function in n variables, $\mathbf{x} \in \mathbb{R}^n$, and there exist m linear equality constraints given by the matrix equation $C\mathbf{x} = \mathbf{d}$. For a solution to exist we must have $m < n$. Thus we have:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T \mathbf{x} \\ \text{subject to:} \quad & C\mathbf{x} = \mathbf{d} \end{aligned}$$

where A is an $n \times n$ positive definite matrix, and C is an $m \times n$ matrix with $\text{rank } C = m$.

The Lagrangian is easily formed as

$$L(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} + \mathbf{b}^T \mathbf{x} + \lambda^T [C\mathbf{x} - \mathbf{d}]$$

where λ is a column vector of m Lagrange multipliers. The necessary conditions for an extremum are now written as:

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}^*, \lambda^*) = A\mathbf{x}^* + \mathbf{b} + C^T \lambda^* = \mathbf{0} \\ \nabla_{\lambda} L(\mathbf{x}^*, \lambda^*) = C\mathbf{x}^* - \mathbf{d} = \mathbf{0} \end{cases}$$

which are a system of $n + m$ equations and $n + m$ unknowns.

These can be succinctly written as:

$$\begin{bmatrix} A & C^T \\ C & O \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -\mathbf{b} \\ \mathbf{d} \end{bmatrix}$$

If we define a new matrix $M \in \mathbb{R}^{(n+m) \times (n+m)}$ such that

$$M = \begin{bmatrix} A & C^T \\ C & O \end{bmatrix}$$

then we can write the solution as

If M is nonsingular, then M^{-1} exists, and the solution $(\mathbf{x}^*, \lambda^*)$ exists.

Alternatively, we can develop a solution by first solving for \mathbf{x}^* in terms of λ^* from the first set of equations. Therefore we have

$$\mathbf{x}^* = -(A^{-1})(\mathbf{b} + C^T \lambda^*) = -A^{-1} \mathbf{b} - A^{-1} C^T \lambda^*$$

and substituting this into the second set of equations

$$C(-A^{-1} \mathbf{b} - A^{-1} C^T \lambda^*) = \mathbf{d}.$$

This is easily solved for the Lagrange multipliers as

$$\lambda^* = (-CA^{-1}C^T)^{-1}(\mathbf{d} + CA^{-1}\mathbf{b}).$$

This is substituted back into the solution $\mathbf{x}^* = -A^{-1}\mathbf{b} - A^{-1}C^T\lambda^*$ which gives

$$\mathbf{x}^* = -A^{-1}\mathbf{b} - A^{-1}C^T((-CA^{-1}C^T)^{-1}(\mathbf{d} + CA^{-1}\mathbf{b}))$$

Example:

As an example, consider the minimization of the linearly constrained positive definite quadratic function

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 \\ & x_1 + 2x_2 + 3x_3 = 7 \\ \text{subject to:} \quad & 2x_1 + 2x_2 + x_3 = \frac{9}{2} \end{aligned}$$

In matrix notation, the objective function becomes

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T I \mathbf{x}$$

$$A = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and the constraints are written as

We can now form the *Lagrangian* as

$$L(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^T\mathbf{x} + \lambda^T[C\mathbf{x} - d]$$

with the necessary conditions for an extremum given by

$$\begin{bmatrix} I & C^T \\ C & O \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 3 & 1 \\ 1 & 2 & 3 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \\ \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 7 \\ \frac{9}{2} \end{bmatrix}$$

We could invert the matrix

$$M = \begin{bmatrix} I & C^T \\ C & O \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 2 & 2 \\ 0 & 0 & 1 & 3 & 1 \\ 1 & 2 & 3 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 \end{bmatrix}$$

in order to solve for the $[\mathbf{x}^* \ \lambda^*]^T$ variables. On the other hand, we can proceed as we did previously by first solving for \mathbf{x}^* in terms of λ^* from the first set of equations. Thus we have

$$\mathbf{x}^* + C^T\lambda^* = \mathbf{0}$$

$$\mathbf{x}^* = -C^T\lambda^*$$

which can be substituted into the second set of equations

$$C(-C^T\lambda^*) = d$$

This can be solved for λ^* as

$$\lambda^* = -(CC^T)^{-1}d$$

For our particular problem, we have

Inverting this we have

$$(CC^T)^{-1} = \frac{1}{45} \begin{bmatrix} 9 & -9 \\ -9 & 14 \end{bmatrix}$$

Therefore we have the Lagrange multipliers

$$\lambda^* = \frac{-1}{45} \begin{bmatrix} 9 & -9 \\ -9 & 14 \end{bmatrix} \begin{bmatrix} 7 \\ 9/2 \end{bmatrix} = \frac{-1}{45} \begin{bmatrix} 22.5 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \end{bmatrix}$$

From the second equation we have the solution we were looking for

$$\mathbf{x}^* = -C^T \lambda^* = - \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -1/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ 3/2 \end{bmatrix}$$

and the minimum of the function is $f(\mathbf{x}^*) = 3.5$.

7.3 Example: Economical Dispatch of Power

Consider the problem of economically dispatching power from n power plants to m loads over a transmission network. We wish to determine the most economical way to generate power so that the demand is satisfied. Let P_i denote the power produced by plant i , $F_i(P_i)$ be the cost of generating power P_i in the plant i , and let the function $\mathcal{L}(P_1, \dots, P_n) = \mathcal{L}(\mathbf{P})$ represent the power lost due to transmission over the lossy transmission lines. The total demand for power is denoted as D . The objective is to minimize the cost of producing power and at the same time meet all the demand. Thus we have

$$\underset{P_i}{\text{minimize}} \quad \sum_{i=1}^n F_i(P_i)$$

subject to the demand that all the demand be met which can be written as

where $\sum_{i=1}^n P_i$ is the total power produced. We can now form the *Lagrangian* as:

$$L(\mathbf{P}, \lambda) = \sum_{i=1}^n F_i(P_i) + \lambda \left[\sum_{i=1}^n P_i - \mathcal{L}(\mathbf{P}) - D \right]$$

The necessary conditions can be written as

$$\frac{\partial L}{\partial P_i} = \frac{dF_i}{dP_i} + \lambda - \lambda \frac{\partial \mathcal{L}}{\partial P_i} = 0 \quad i = 1, 2, \dots, n \quad (7.6)$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n P_i - \mathcal{L}(\mathbf{P}) - D = 0 \quad (7.7)$$

From the first of these we have

$$\frac{dF_i}{dP_i} = \lambda \left(\frac{\partial \mathcal{L}}{\partial P_i} - 1 \right) = \lambda \mathcal{L}_i$$

where we define the *loss factor* as

$$\mathcal{L}_i = \frac{\partial \mathcal{L}}{\partial P_i} - 1$$

Therefore, we have

$$\frac{1}{\mathcal{L}_i} \frac{dF_i}{dP_i} = \lambda$$

and λ is determined by solving (7.6) and (7.7) simultaneously.

7.4 Economic Interpretation of Lagrange Multipliers

Consider

$$\text{minimize}_{\mathbf{x}} \quad F(\mathbf{x})$$

$$\text{subject to:} \quad g_i(\mathbf{x}) = b_i \quad i = 1, 2, \dots, m$$

and think of the values b_i as some available resource. Now we wish to investigate how the solution $(\mathbf{x}^*, \lambda^*)$ varies as the b_i vary. We first write the Lagrangian as

and thinking of this as a function of \mathbf{b} we can write it more explicitly as

$$L(\mathbf{b}) = F(\mathbf{x}(\mathbf{b})) + \lambda^T(\mathbf{b})[\mathbf{g}(\mathbf{x}(\mathbf{b})) - \mathbf{b}]$$

and differentiating this with respect to b_i

$$\begin{aligned} \frac{\partial L}{\partial b_i} &= \sum_{j=1}^n \frac{\partial F}{\partial x_j} \frac{\partial x_j}{\partial b_i} + \lambda^T(\mathbf{b}) \sum_{j=1}^n \frac{\partial \mathbf{g}}{\partial x_j} \frac{\partial x_j}{\partial b_i} - \lambda_i + \frac{\partial}{\partial b_i} \lambda^T(\mathbf{b})[\mathbf{g}(\mathbf{x}(\mathbf{b})) - \mathbf{b}] \\ &= \sum_{j=1}^n \frac{\partial F}{\partial x_j} \frac{\partial x_j}{\partial b_i} + \sum_{k=1}^m \sum_{j=1}^n \lambda_k \frac{\partial g_k}{\partial x_j} \frac{\partial x_j}{\partial b_i} - \lambda_i + \sum_{k=1}^m [g_k(\mathbf{x}) - b_k] \frac{\partial \lambda_k}{\partial b_i} \\ &= \sum_{j=1}^n \frac{\partial x_j}{\partial b_i} \left[\frac{\partial F}{\partial x_j} + \sum_{k=1}^m \lambda_k \frac{\partial g_k}{\partial x_j} \right] + \sum_{k=1}^m [g_k - b_k] \frac{\partial \lambda_k}{\partial b_i} - \lambda_i \end{aligned}$$

At $(\mathbf{x}^*, \lambda^*)$ the necessary conditions for an extremum give:

$$\frac{\partial F}{\partial x_i} + \sum_{k=1}^m \lambda_k \frac{\partial g_k}{\partial x_j} = \frac{\partial L}{\partial x_j}(\mathbf{x}^*, \lambda^*) = 0$$

and

$$(g_k - b_k)|_{\mathbf{x}^*} = 0$$

and therefore

$$\frac{\partial L}{\partial b_i}(\mathbf{x}^*, \lambda^*) = -\lambda_i^* = \frac{\partial F}{\partial b_i} \Big|_{\mathbf{x}^*}, \quad i = 1, 2, \dots, m$$

since $F(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*)$

Therefore, the Lagrange multiplier λ_i^* gives the rate of change of the objective function with respect to the resource b_i , sometimes called “*shadow prices*” or “*sensitivity coefficients*”.

Example: optimal allocation of a scarce resource between two processes. Say the total amount of a resource is b . Process 1 gives us a return based on the amount of resource b used of $F(x_1)$, while process 2 gives us a return of $F(x_2)$.

$$\underset{\mathbf{x}}{\text{maximize}} \quad F = F(x_1) + F(x_2)$$

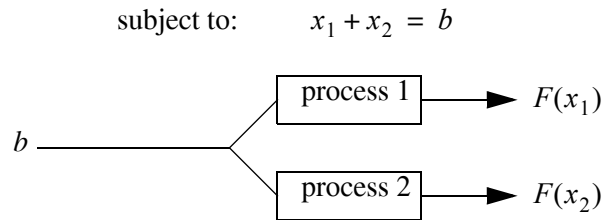


Figure 7.2 Two processes with a single resource.

As a particular example, consider

$$F_1(x_1) = 50 - (x_1 - 2)^2$$

$$F_2(x_2) = 50 - (x_2 - 2)^2$$

the Lagrangian

$$L(\mathbf{x}, \lambda) = 100 - (x_1 - 2)^2 - (x_2 - 2)^2 + \lambda[x_1 + x_2 - b]$$

and the necessary conditions are

$$\begin{cases} \frac{\partial L}{\partial x_1} = -2(x_1^* - 2) + \lambda^* = 0 \\ \frac{\partial L}{\partial x_2} = -2(x_2^* - 2) + \lambda^* = 0 \\ \frac{\partial L}{\partial \lambda} = x_1^* + x_2^* - b = 0 \end{cases}$$

solving we have

$$(x_1^* - 2) - (x_2^* - 2) = 0$$

$$x_1^* + x_2^* - b = 0$$

$$2x_1^* = b$$

$$\begin{cases} x_1^* = b/2 \\ x_2^* = b/2 \\ \lambda^* = b - 4 \end{cases}$$

The sensitivity can now be calculated as

therefore if $b < 4$ this implies that increasing b will increase the return; if $b > 4$ then increasing b means a decreasing return; if $b = 4$ then $\lambda^* = 0$ which means that the solutions to the constrained and unconstrained problems are the same.

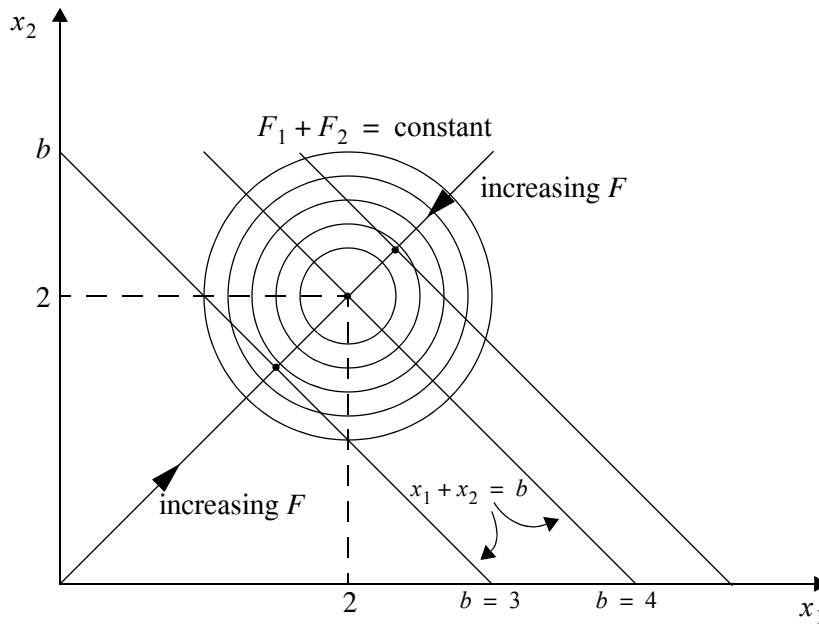


Figure 7.3 Contour plot of F in the example with linear equality constraint.

We can get a clearer view of what is going on by viewing the contour plot shown in Figure 7.3. The circles represent level curves for $F = F(x_1) + F(x_2)$. The straight lines represent the single inequality constraint for different values of b . We see that when $b = 4$ the constraint line passes through centre of circles, that is at point $(2, 2)$. Here $\lambda^* = 0$ and the constrained and unconstrained problems have the same solution.

When $b < 4$ increasing b results in an increasing value for the maximum of $F_1 + F_2$

$$(-\lambda^*) = 4 - b > 0$$

when $b > 4$ increasing b results in an decreasing value for the maximum of $F_1 + F_2$

$$(-\lambda^*) = 4 - b < 0$$

What do the Lagrange multipliers tell us about the purchase of more resource b ? We know that the rate of change of the profit is $-\lambda^*$ with respect to increasing the resource b . If the cost of b is y [\$/unit of b] then as long as $-\lambda^* > y$ we should purchase more b . For this reason, $-\lambda^*$ is sometimes called the marginal return. In the present example $-\lambda^* = 4 - b$ (in general λ will depend on b). If the cost of buying more b is $y = 1$ \$/unit, then we should buy enough to bring $b = 3$ at which point $y = 1 = -\lambda^*$. Increasing b further will cost more than the increased profit it will produce.