Constrained Optimization Using Lagrange Multipliers

CEE 201L. Uncertainty, Design, and Optimization
Department of Civil and Environmental Engineering
Duke University
Henri P. Gavin and Jeffrey T. Scruggs
Spring, 2016

In optimal design problems, values for a set of \( n \) design variables, \((x_1, x_2, \cdots, x_n)\), are to be found that minimize a scalar-valued objective function of the design variables, such that a set of \( m \) inequality constraints, are satisfied. Constrained optimization problems are generally expressed as

\[
\min_{x_1, x_2, \cdots, x_n} J = f(x_1, x_2, \cdots, x_n) \quad \text{such that} \quad g_1(x_1, x_2, \cdots, x_n) \leq 0 \\
g_2(x_1, x_2, \cdots, x_n) \leq 0 \\
\vdots \\
g_m(x_1, x_2, \cdots, x_n) \leq 0
\]  

(1)

If the objective function is quadratic in the design variables and the constraint equations are linear in the design variables, the optimization problem usually has a unique solution.

Consider the simplest constrained minimization problem:

\[
\min_x \frac{1}{2} kx^2 \quad \text{such that} \quad x \geq b .
\]  

(2)

This problem has a single design variable, the objective function is quadratic \((J = \frac{1}{2} kx^2)\), there is a single constraint inequality, and it is linear in \( x \) \((g(x) = b - x)\). If \( g > 0 \), the constraint equation constrains the optimum and the optimal solution, \( x^* \), is given by \( x^* = b \). If \( g \leq 0 \), the constraint equation does not constrain the optimum and the optimal solution is given by \( x^* = 0 \). Not all optimization problems are so easy; most optimization methods require more advanced methods. The methods of Lagrange multipliers is one such method, and will be applied to this simple problem.

Lagrange multiplier methods involve the modification of the objective function through the addition of terms that describe the constraints. The objective function \( J = f(x) \) is augmented by the constraint equations through a set of non-negative multiplicative Lagrange multipliers, \( \lambda_j \geq 0 \). The augmented objective function, \( J_A(x) \), is a function of the \( n \) design variables and \( m \) Lagrange multipliers,

\[
J_A(x_1, x_2, \cdots, x_n, \lambda_1, \lambda_2, \cdots, \lambda_m) = f(x_1, x_2, \cdots, x_n) + \sum_{j=1}^{m} \lambda_j g_j(x_1, x_2, \cdots, x_n)
\]  

(3)

For the problem of equation (2), \( n = 1 \) and \( m = 1 \), so

\[
J_A(x, \lambda) = \frac{1}{2} kx^2 + \lambda (b - x)
\]  

(4)

The Lagrange multiplier, \( \lambda \), serves the purpose of modifying (augmenting) the objective function from one quadratic \((\frac{1}{2} kx^2)\) to another quadratic \((\frac{1}{2} kx^2 - \lambda x + \lambda b)\) so that the minimum of the modified quadratic satisfies the constraint \((x \geq b)\).
Case 1: $b = 1$

If $b = 1$ then the minimum of $\frac{1}{2}kx^2$ is constrained by the inequality $x \geq b$, and the optimal value of $\lambda$ should minimize $J_A(x, \lambda)$ at $x = b$. Figure 1(a) plots $J_A(x, \lambda)$ for a few non-negative values of $\lambda$ and Figure 1(b) plots contours of $J_A(x, \lambda)$.

$$J_A(x, \lambda) = \frac{1}{2}kx^2 - \lambda x + \lambda b$$ for $b = 1$ and $k = 2$. $(x^*, \lambda^*) = (1, 2)$
Figure 1 shows that:

- $J_A(x, \lambda)$ is independent of $\lambda$ at $x = b$,
- $J_A(x, \lambda)$ is minimized at $x^* = b$ for $\lambda^* = 2$,
- the surface $J_A(x, \lambda)$ is a saddle shape,
- the point $(x^*, \lambda^*) = (1, 2)$ is a saddle point,
- $J_A(x^*, \lambda) \leq J_A(x^*, \lambda^*) \leq J_A(x, \lambda^*)$,
- $\min_x J_A(x, \lambda^*) = \max_\lambda J_A(x^*, \lambda) = J_A(x^*, \lambda^*)$

Saddle points have no slope.

\[
\frac{\partial J_A(x, \lambda)}{\partial x} \bigg|_{x=x^*, \lambda=\lambda^*} = 0 \quad (5a)
\]
\[
\frac{\partial J_A(x, \lambda)}{\partial \lambda} \bigg|_{x=x^*, \lambda=\lambda^*} = 0 \quad (5b)
\]

For this problem,

\[
\frac{\partial J_A(x, \lambda)}{\partial x} \bigg|_{x=x^*, \lambda=\lambda^*} = 0 \Rightarrow \ kx^* - \lambda^* = 0 \Rightarrow \lambda^* = kx^* \quad (6a)
\]
\[
\frac{\partial J_A(x, \lambda)}{\partial \lambda} \bigg|_{x=x^*, \lambda=\lambda^*} = 0 \Rightarrow -x^* + b = 0 \Rightarrow x^* = b \quad (6b)
\]

This example has a physical interpretation. The objective function $J = \frac{1}{2}kx^2$ represents the potential energy in a spring. The minimum potential energy in a spring corresponds to a stretch of zero ($x^* = 0$). The constrained problem:

\[
\min_x \frac{1}{2}kx^2 \quad \text{such that} \quad x \geq 1
\]

means “minimize the potential energy in the spring such that the stretch in the spring is greater than or equal to 1.” The solution to this problem is to set the stretch in the spring equal to the smallest allowable value ($x^* = 1$). The force applied to the spring in order to achieve this objective is $f = kx^*$. This force is the Lagrange multiplier for this problem, ($\lambda^* = kx^*$).

The Lagrange multiplier is the force required to enforce the constraint.
Case 2: $b = -1$

If $b = -1$ then the minimum of $\frac{1}{2}kx^2$ is not constrained by the inequality $x \geq b$. The derivation above would give $x^* = -1$, with $\lambda^* = -k$. The negative value of $\lambda^*$ indicates that the constraint does not affect the optimal solution, and $\lambda^*$ should therefore be set to zero. Setting $\lambda^* = 0$, $J_A(x, \lambda)$ is minimized at $x^* = 0$. Figure 2(a) plots $J_A(x, \lambda)$ for a few negative values of $\lambda$ and Figure 2(b) plots contours of $J_A(x, \lambda)$.

![Figure 2](image_url)

**Figure 2.** $J_A(x, \lambda) = \frac{1}{2}kx^2 - \lambda x + \lambda b$ for $b = -1$ and $k = 2$. $(x^*, \lambda^*) = (0, 0)$
Figure 2 shows that:

• $J_A(x, \lambda)$ is independent of $\lambda$ at $x = b$,

• the saddle point of $J_A(x, \lambda)$ occurs at a negative value of $\lambda$, so $\partial J_A/\partial \lambda \neq 0$ for any $\lambda \geq 0$.

These examples illustrate the properties of Lagrange-multiplier problems.

If the inequality $g(x) \leq 0$ constrains $\min f(x)$ then the optimum point of the augmented objective $J_A(x, \lambda) = f(x) + \lambda g(x)$ is minimized with respect to $x$ and maximized with respect to $\lambda$. The optimization problem of equation (1) may be written in terms of the augmented objective function (equation (3)),

$$\max_{\lambda_1, \lambda_2, \ldots, \lambda_m, x_1, x_2, \ldots, x_n} \min J_A(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m) \quad \text{such that} \quad \lambda_j \geq 0 \; \forall \; j \quad (7)$$

The conditions

$$\frac{\partial J_A}{\partial x_k} \bigg|_{x_i = x_i^*, \lambda_j = \lambda_j^*} = 0 \quad (8a)$$

$$\frac{\partial J_A}{\partial \lambda_k} \bigg|_{x_i = x_i^*, \lambda_j = \lambda_j^*} = 0 \quad (8b)$$

define the optimal design variables $x_i^*$ such that $g_j(x_1, x_2, \ldots, x_n) \leq 0$.

If an inequality $g_j(x_1, x_2, \ldots, x_n)$ constrains the optimum point, the corresponding Lagrange multiplier, $\lambda_j$ is positive. Lagrange multipliers for non-active constraints would be negative if they were not constrained to be non-negative.

The value of the Lagrange multiplier at the optimum point is the sensitivity of the objective function $J = f(x)$ to changes in the constraint value, $b$. Changing the constraint level from $b$ to $b + \delta b$ changes the constrained objective function from $J = f(x^*)$ to $J = f(x^*) + \lambda^* \delta b$.

---

Figure 3. If increasing the constraint, $g(x)$, results in an improved objective (a reduced cost), then $\lambda < 0$, and the constraint $g(x) \leq 0$ should not bind the optimum point. If increasing the constraint results in an increased cost, then $\lambda > 0$, and the constraint $g(x) \leq 0$ must be enforced at the optimum point.
Multivariable Quadratic Programming

Quadratic optimization problems with a single design variable and a single linear inequality constraint are easy enough. In problems with many design variables \((n \gg 1)\) and many inequality constraints \((m \gg 1)\), determining which inequality constraints are enforced at the optimum point can be difficult. Numerical methods used in solving such problems involve iterative trial-and-error approaches to find the set of “active” inequality constraints.

Consider a second example with \(n = 2\) and \(m = 2\):

\[
\begin{align*}
\min_{x_1, x_2} & \quad x_1^2 + 0.5x_1 + 3x_1x_2 + 5x_2^2 \\
\text{such that} & \quad 3x_1 + 2x_2 + 2 \leq 0 \\
& \quad 15x_1 - 3x_2 - 1 \leq 0
\end{align*}
\]

This example also has a quadratic objective function and inequality constraints that are linear in the design variables. Contours of the objective function and the two inequality constraints are plotted in Figure 4. The feasible region of these two inequality constraints is to the left of the lines in the figure and are labeled as “\(g_1 \text{ ok}\)” and “\(g_2 \text{ ok}\)”. This figure shows that the inequality \(g_1(x_1, x_2)\) constrains the solution and the inequality \(g_2(x_1, x_2)\) does not. This is visible in Figure 4 with \(n = 2\), but for more complicated problems it may not be immediately clear which inequality constraints are “active.”

Using the method of Lagrange multipliers, the augmented objective function is

\[
J_A(x_1, x_2, \lambda_1, \lambda_2) = x_1^2 + 0.5x_1 + 3x_1x_2 + 5x_2^2 + \lambda_1(3x_1 + 2x_2 + 2) + \lambda_2(15x_1 - 3x_2 - 1)
\]

Unlike the first examples with \(n = 1\) and \(m = 1\), we cannot plot contours of \(J_A(x_1, x_2, \lambda_1, \lambda_2)\) since this would be a plot in four-dimensional space. Nonetheless, the same optimality conditions hold.

\[
\begin{align*}
\min_{x_1} J_A & \Rightarrow \frac{\partial J_A}{\partial x_1} \bigg|_{x_1^*, x_2^*, \lambda_1^*, \lambda_2^*} = 0 \Rightarrow 2x_1^* + 0.5 + 3x_2^* + 3\lambda_1^* + 15\lambda_2^* = 0 \\
\min_{x_2} J_A & \Rightarrow \frac{\partial J_A}{\partial x_2} \bigg|_{x_1^*, x_2^*, \lambda_1^*, \lambda_2^*} = 0 \Rightarrow 3x_1^* + 10x_2^* + 2\lambda_1^* - 3\lambda_2^* = 0 \\
\max_{\lambda_1} J_A & \Rightarrow \frac{\partial J_A}{\partial \lambda_1} \bigg|_{x_1^*, x_2^*, \lambda_1^*, \lambda_2^*} = 0 \Rightarrow 3x_1^* + 2x_2^* + 2 = 0 \\
\max_{\lambda_2} J_A & \Rightarrow \frac{\partial J_A}{\partial \lambda_2} \bigg|_{x_1^*, x_2^*, \lambda_1^*, \lambda_2^*} = 0 \Rightarrow 15x_1^* - 3x_2^* - 1 = 0
\end{align*}
\]

If the objective function is quadratic in the design variables and the constraints are linear in the design variables, the optimality conditions are simply a set of linear equations in the design variables and the Lagrange multipliers. In this example the optimality conditions are expressed as four linear equations with four unknowns. In general we may not know which inequality constraints are active. If there are only a few constraint equations it’s not too hard to try all combinations of any number of constraints, fixing the Lagrange multipliers for the other inequality constraints equal to zero.
Let’s try this now!

• First, let’s find the unconstrained minimum by assuming neither constraint \( g_1(x_1, x_2) \) or \( g_2(x_1, x_2) \) is active, \( \lambda_1^* = 0, \lambda_2^* = 0, \) and

\[
\begin{bmatrix}
2 & 3 & 3 & 15 \\
3 & 10 & 2 & -3 \\
3 & 2 & 0 & 0 \\
15 & -3 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_1^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.5 \\
0 \\
-2 \\
1 \\
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_1^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.45 \\
-0.14 \\
-2 \\
1 \\
\end{bmatrix},
\] (12)

which is the unconstrained minimum shown in Figure 4 as a “∗”. Plugging this solution into the constraint equations gives \( g_1(x_1^*, x_2^*) = 0.93 \) and \( g_2(x_1^*, x_2^*) = -8.17, \) so the unconstrained minimum is not feasible with respect to constraint \( g_1, \) since \( g_1(-0.45, 0.14) > 0. \)

• Next, assuming both constraints \( g_1(x_1, x_2) \) and \( g_2(x_1, x_2) \) are active, optimal values for \( x_1^*, x_2^*, \lambda_1^*, \) and \( \lambda_2^* \) are sought, and all four equations must be solved together.

\[
\begin{bmatrix}
2 & 3 & 3 & 15 \\
3 & 10 & 2 & -3 \\
3 & 2 & 0 & 0 \\
15 & -3 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_1^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.5 \\
0 \\
-2 \\
1 \\
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_1^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.10 \\
-0.85 \\
3.55 \\
-0.56 \\
\end{bmatrix},
\] (13)

which is the constrained minimum shown in Figure 4 as a “o” at the intersection of the \( g_1 \) line with the \( g_2 \) line in Figure 4. Note that \( \lambda_2^* < 0 \) indicating that constraint \( g_2 \) is not active: \( g_2(-0.10, 0.85) = 0 \) (ok). Enforcing the constraint \( g_2 \) needlessly compromises the optimality of this solution. So, while this solution is feasible (both \( g_1 \) and \( g_2 \) evaluate to zero), the solution could be improved by letting go of the \( g_2 \) constraint and moving along the \( g_1 \) line.

• So, assuming only constraint \( g_1 \) is active, \( g_2 \) is not active, \( \lambda_2^* = 0, \) and

\[
\begin{bmatrix}
2 & 3 & 3 \\
3 & 10 & 2 \\
3 & 2 & 0 \\
15 & -3 & 0 \\
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_1^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.5 \\
0 \\
-2 \\
1 \\
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_1^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.81 \\
0.21 \\
0.16 \\
\end{bmatrix},
\] (14)

which is the constrained minimum as a “o” on the \( g_1 \) line in Figure 4. Note that \( \lambda_1^* > 0, \) which indicates that this constraint is active. Plugging \( x_1^* \) and \( x_2^* \) into \( g_2(x_1, x_2) \) gives a value of -13.78, so this constrained minimum is feasible with respect to both constraints. This is the solution we’re looking for.

• As a final check, assuming only constraint \( g_2 \) is active, \( g_1 \) is not active, \( \lambda_1^* = 0, \) and

\[
\begin{bmatrix}
2 & 3 & 15 \\
3 & 10 & -3 \\
15 & -3 & 0 \\
\end{bmatrix}
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
-0.5 \\
0 \\
1 \\
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
x_1^* \\
x_2^* \\
\lambda_2^* \\
\end{bmatrix}
= \begin{bmatrix}
0.06 \\
-0.03 \\
-0.04 \\
\end{bmatrix},
\] (15)

which is the constrained minimum shown in as a “o” on the \( g_2 \) line in Figure 4. Note that \( \lambda_2^* < 0, \) which indicates that this constraint is not active, contradicting our assumption. Further, plugging \( x_1^* \) and \( x_2^* \) into \( g_1(x_1, x_2) \) gives a value of +2.24, so this constrained minimum is not feasible with respect to \( g_1, \) since \( g_1(0.06, -0.03) > 0. \)
Figure 4. Contours of the objective function and the constraint equations for the example of equation (9). (a): $J = f(x_1, x_2)$; (b): $J_A = f(x_1, x_2) + \lambda^* g_1(x_1, x_2)$. Note that the contours of $J_A$ are shifted so that the minimum of $J_A$ is at the optimal point along the $g_1$ line.
Primal and Dual

If it is possible to solve for the design variables in terms of the Lagrange multipliers, then the design variables can be eliminated from the problem and the optimization is simply a maximization over the set of Lagrange multipliers.

Consider the minimization of a quadratic objective function subject to a set of linear inequality constraints

$$\min_{x_1, x_2, \ldots, x_n} \left[ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i H_{ij} x_j \right] \quad \text{such that} \quad \sum_{j=1}^{n} A_{1j} x_j - b_1 \leq 0$$
$$\sum_{j=1}^{n} A_{2j} x_j - b_2 \leq 0$$
$$\vdots$$
$$\sum_{j=1}^{n} A_{mj} x_j - b_m \leq 0$$

(16)

where $H_{ij} = H_{ji}$ and $\sum_i \sum_j x_i H_{ij} x_j > 0$ for any set of design variables. In matrix-vector notation, equation (16) is written

$$\min_{x} \frac{1}{2} x^T H x \quad \text{such that} \quad A x - b \leq 0$$

(17)

The augmented optimization problem is

$$\max_{\lambda_1, \ldots, \lambda_m} \min_{x_1, x_2, \ldots, x_n} \left[ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i H_{ij} x_j + \sum_{k=1}^{m} \lambda_k \left( \sum_{j=1}^{n} A_{kj} x_j - b_k \right) \right]$$

(18)

such that $\lambda_k \geq 0$. Or, in matrix-vector notation,

$$\max_{\lambda} \min_{x} \left[ \frac{1}{2} x^T H x + \lambda^T (A x - b) \right] \quad \text{such that} \quad \lambda \geq 0.$$  

(19)

For this kind of problem, the condition

$$\frac{\partial J_A(x, \lambda)}{\partial x} = 0^T$$

(20)

results in $H x + A^T \lambda = 0$ from which $x = -H^{-1} A^T \lambda$. Substituting this solution into $J_A$ results in

$$\max_{\lambda} \left[ -\frac{1}{2} \lambda^T A H^{-1} A^T \lambda - \lambda^T b \right] \quad \text{such that} \quad \lambda \geq 0,$$

(21)

or

$$\min_{\lambda} \left[ -\frac{1}{2} \lambda^T A H^{-1} A^T \lambda + b^T \lambda \right] \quad \text{such that} \quad \lambda \geq 0,$$

(22)

which is independent of $x$.

Equation (19) is called the primal quadratic programming problem and equation (22) is called the dual quadratic programming problem. The primal problem has $n + m$ unknown variables ($x$ and $\lambda$) whereas the dual problem has only $m$ unknown variables ($\lambda$) and is therefore easier to solve.