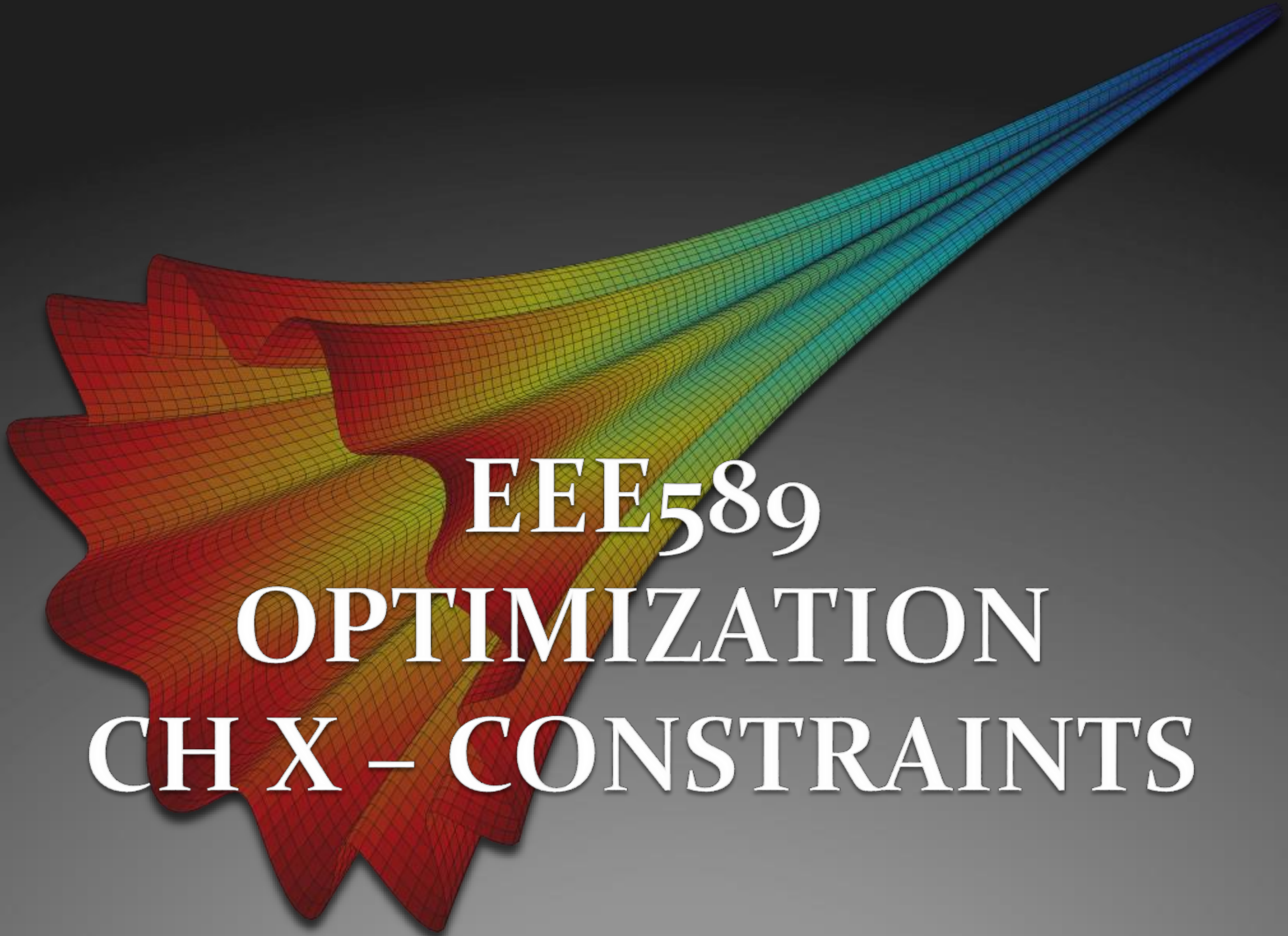


EEE589
OPTIMIZATION
CH X – CONSTRAINTS

Constraints

- Previous chapters have focused on unconstrained problems where the domain of each design variable is the space of real numbers.
- Many problems are constrained, which forces design points to satisfy certain conditions.
- This lecture presents a variety of approaches for transforming problems with constraints into problems without constraints, thereby permitting the use of the optimization algorithms we have already discussed.
- Analytical methods are also discussed, including the concepts of duality and the necessary conditions for optimality under constrained optimization.

Constrained Optimization

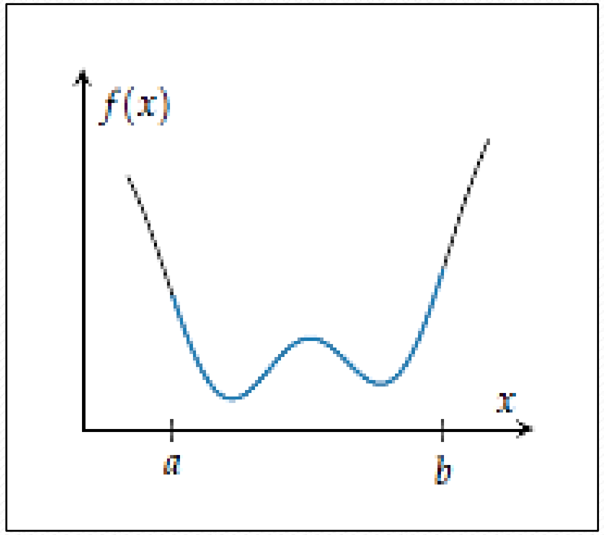
- Minimizing an objective subject to design point restrictions called constraints
- A variety of techniques transform constrained optimization problems into unconstrained problems
- New optimization problem statement

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X} \end{aligned}$$

- The set $\mathcal{X}$ is called the feasible set. In constrained problems, the feasible set is some subset thereof.

Constrained Optimization

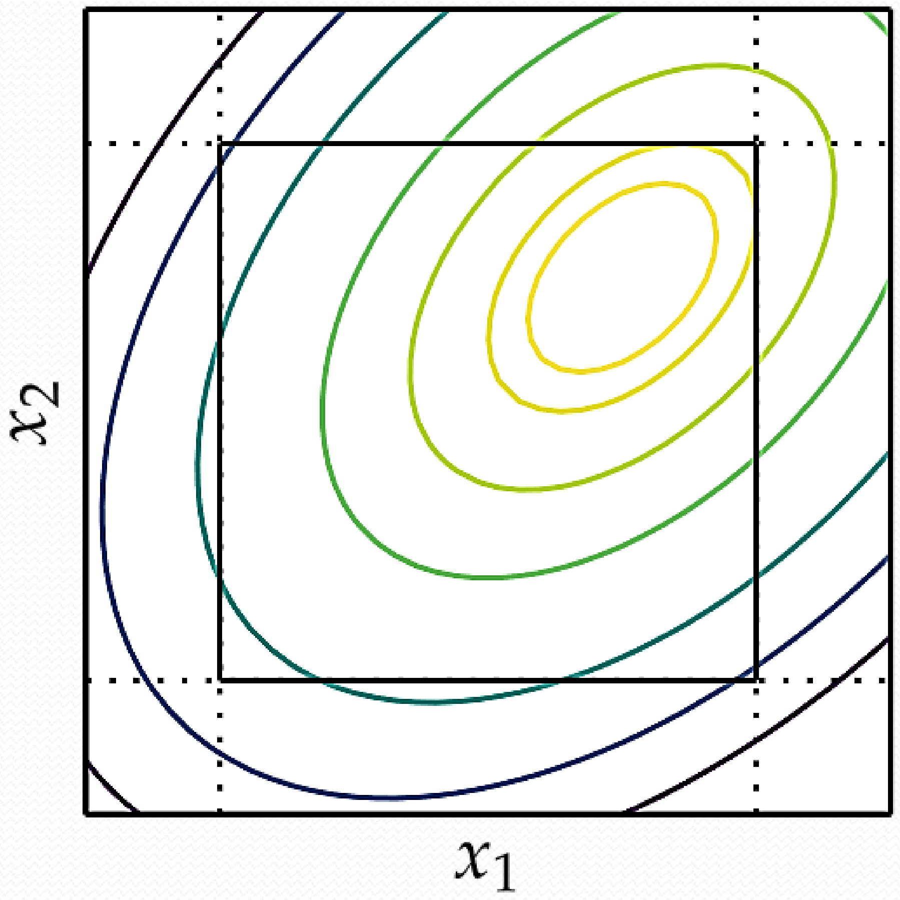
- Some constraints are simply upper or lower bounds on the design variables, as we have seen in bracketed line search, in which x must lie between a and b.
- A bracketing constraint x ∈ [a, b] can be replaced by two inequality constraints: a ≤ x and x ≤ b as shown in figure.



$$\begin{aligned} &\underset{x}{\text{minimize}} && f(x) \\ &\text{subject to} && x \in [a, b] \end{aligned}$$

Constrained Optimization

- In multivariate problems, bracketing the input variables forces them to lie within a hyperrectangle as shown in figure.

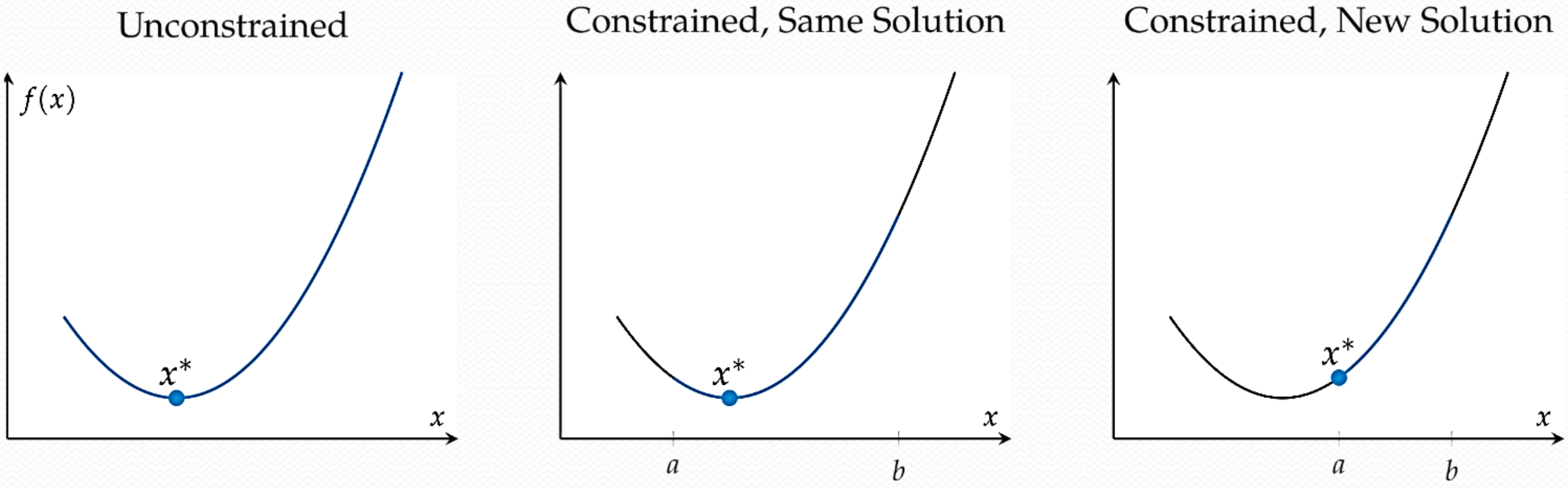


Constrained Optimization

- Constraints arise naturally when formulating real problems.
- A hedge fund manager cannot sell more stock than they have, an airplane cannot have wings with zero thickness, and the number of hours you spend per week on your homework cannot exceed 168.
- We include constraints in such problems to prevent the optimization algorithm from suggesting an infeasible solution.

Constrained Optimization

- Applying constraints to a problem can affect the solution, but this need not be the case as shown in figure.



Constraint Types

- Generally, constraints are formulated using two types

1. Equality constraints: $h(\mathbf{x}) = 0$
2. Inequality constraints: $g(\mathbf{x}) \leq 0$

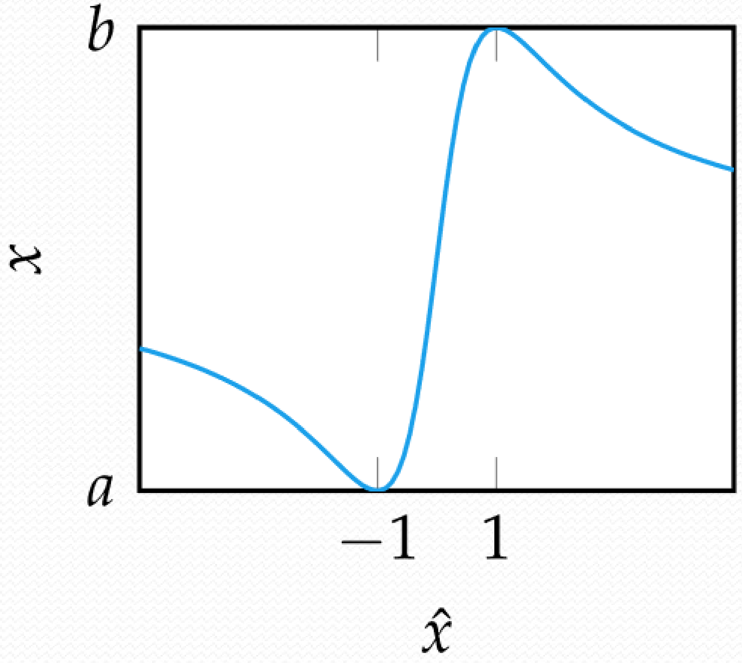
- Any optimization problem can be written as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0 \ \text{ for all } \ i \text{ in } \{1, \ldots, \ell\} \\ & g_j(\mathbf{x}) \leq 0 \ \text{ for all } \ j \text{ in } \{1, \ldots, m\} \end{aligned}$$

Transformations to Remove Constraints

- If necessary, some problems can be reformulated to incorporate constraints into the objective function
- If x is constrained between a and b

$$x = t_{a,b}(\hat{x}) = \frac{b+a}{2} + \frac{b-a}{2}\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)$$



Example demonstrates this process.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x\sin(x) \\ \text{subject to} \quad & 2 \le x \le 6 \end{aligned}$$

$$\underset{\hat{x}}{\text{minimize}} \quad t_{2,6}(\hat{x})\sin(t_{2,6}(\hat{x}))$$

$$\underset{\hat{x}}{\text{minimize}} \quad \left(4+2\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)\right)\sin\left(4+2\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)\right)$$

Consider the optimization problem

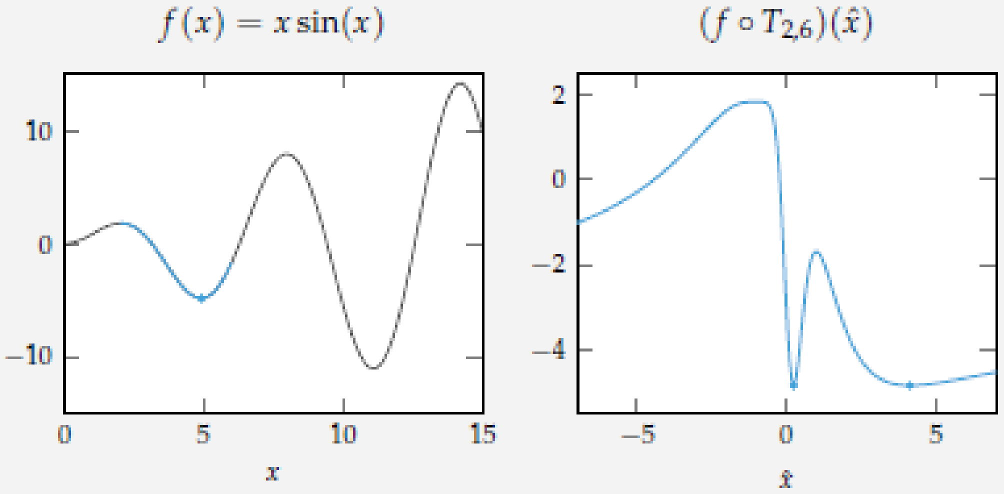
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x\sin(x) \\ \text{subject to} \quad & 2 \le x \le 6 \end{aligned}$$

We can transform the problem to remove the constraints:

$$\underset{\hat{x}}{\text{minimize}} \quad t_{2,6}(\hat{x})\sin(t_{2,6}(\hat{x}))$$

$$\underset{\hat{x}}{\text{minimize}} \quad \left(4+2\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)\right)\sin\left(4+2\left(\frac{2\hat{x}}{1+\hat{x}^2}\right)\right)$$

We can use the optimization method of our choice to solve the unconstrained problem. In doing so, we find two minima: $\hat{x} \approx 0.242$ and $\hat{x} \approx 4.139$, both of which have a function value of approximately -4.814.



The solution for the original problem is obtained by passing $\hat{x}$ through the transform. Both values of $\hat{x}$ produce $x = t_{2,6}(\hat{x}) \approx 4.914$.

Transformations to Remove Constraints

- Some equality constraints can be used to solve for x_n given $x_1, \dots, x_{n-1}$.
- In other words, if we know the first $n-1$ components of $\mathbf{x}$, we can use the constraint equation to obtain x_n.
- In such cases, the optimization problem can be reformulated over $x_1, \dots, x_{n-1}$ instead, removing the constraint and removing one design variable.

Example demonstrates this process.

Consider the constraint:

$$h(\mathbf{x}) = x_1^2 + x_2^2 + \cdots + x_n^2 - 1 = 0$$

We can solve for x_n using the first $n-1$ variables:

$$x_n = \pm\sqrt{1 - x_1^2 + x_2^2 + \cdots + x_{n-1}^2}$$

We can transform

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h(\mathbf{x}) = 0 \end{aligned}$$

into

$$\underset{x_1,\ldots,x_{n-1}}{\text{minimize}} \quad f\left(\left[x_1, \ldots, x_{n-1}, \pm\sqrt{1 - x_1^2 + x_2^2 + \cdots + x_{n-1}^2}\right]\right)$$

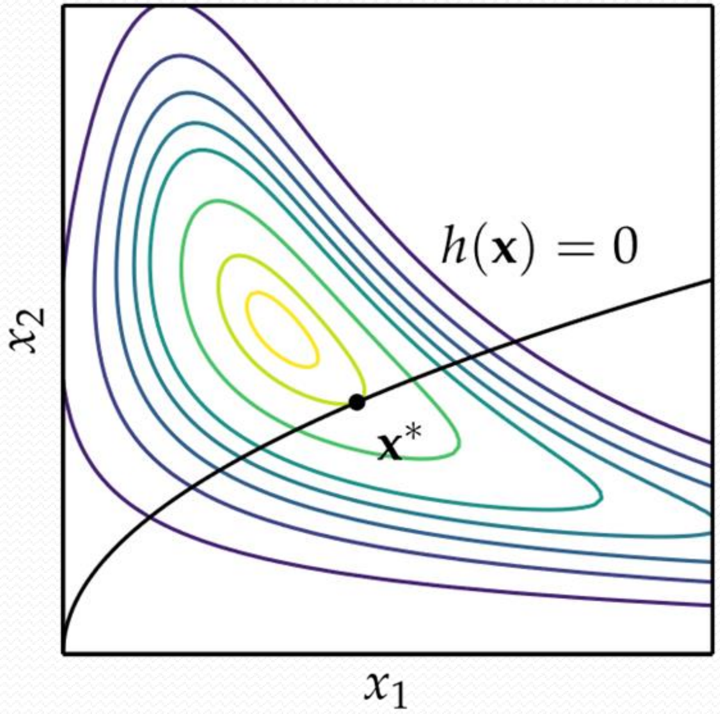
Lagrange Multipliers

- The method of *Lagrange multipliers* is used to optimize a function subject to equality constraints.
- Consider an optimization problem with a single equality constraint:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h(\mathbf{x}) = 0 \end{aligned}$$

where f and h have continuous partial derivatives.

- Example discusses such a problem.
- Intuitively, the method of Lagrange multipliers finds the point x* where the constraint function is orthogonal to the gradient.



Consider the minimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -\exp\left(-\left(x_1 x_2 - \frac{3}{2}\right)^2 - (x_2 - \frac{3}{2})^2\right) \\ \text{subject to} \quad & x_1 - x_2^2 = 0 \end{aligned}$$

We substitute the constraint $x_1 = x_2^2$ into the objective function to obtain an unconstrained objective:

$$f_{\text{unc}} = -\exp\left(-\left(x_2^3 - \frac{3}{2}\right)^2 - \left(x_2 - \frac{3}{2}\right)^2\right)$$

whose derivative is:

$$\frac{\partial}{\partial x_2} f_{\text{unc}} = 6\exp\left(-\left(x_2^3 - \frac{3}{2}\right)^2 - \left(x_2 - \frac{3}{2}\right)^2\right)\left(x_2^5 - \frac{3}{2}x_2^2 + \frac{1}{3}x_2 - \frac{1}{2}\right)$$

Setting the derivative to zero and solving for x_2 yields $x_2 \approx 1.165$. The solution to the original optimization problem is thus $\mathbf{x}^* \approx [1.358, 1.165]$. The optimum lies where the contour line of f is aligned with h.

If the point x^* optimizes f along h, then its directional derivative at x^* along h must be zero. That is, small shifts of x^* along h cannot result in an improvement.

The contour lines of f are lines of constant f. Thus, if a contour line of f is tangent to h, then the directional derivative of h at that point, along the direction of the contour $h(\mathbf{x}) = 0$, must be zero.

Lagrange Multipliers

- The method of Lagrange multipliers is used to compute where a contour line of f is aligned with the contour line of h(x) = 0.
- Since the gradient of a function at a point is perpendicular to the contour line of that function through that point, we know the gradient of h will be perpendicular to the contour line h(x) = 0.
- Hence, we need to find where the gradient of f and the gradient of h are aligned.
- We seek the best x such that the constraint h(x)=0 is satisfied and the gradients are aligned $\nabla f(\mathbf{x}) = \lambda \nabla h(\mathbf{x})$ for some *Lagrange multiplier* λ. We need the scalar λ because the magnitudes of the gradients may not be the same.
- We can formulate the *Lagrangian*, which is a function of the design variables and the multiplier

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda h(\mathbf{x})$$

Lagrange Multipliers

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h(\mathbf{x}) = 0 \end{aligned}$$

1. Form Lagrangian equation

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda h(\mathbf{x})$$

2. Set gradient of Lagrangian with respect to **x** to zero to get

$$\nabla f(\mathbf{x}) = \lambda \nabla h(\mathbf{x})$$

3. Use constraint equation $h(\mathbf{x}) = 0$, solve for **x** and λ

- Example demonstrates this approach.

Consider the minimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -\exp\left(-\left(x_1 x_2 - \frac{3}{2}\right)^2 - (x_2 - \frac{3}{2})^2\right) \\ \text{subject to} \quad & x_1 - x_2^2 = 0 \end{aligned}$$

We can use the method of Lagrange multipliers to solve the problem in example 10.3. We form the Lagrangian

$$\mathcal{L}(x_1, x_2, \lambda) = -\exp\left(-\left(x_1 x_2 - \frac{3}{2}\right)^2 - \left(x_2 - \frac{3}{2}\right)^2\right) - \lambda(x_1 - x_2^2)$$

and compute the gradient

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= 2x_2 f(\mathbf{x})\left(\frac{3}{2} - x_1 x_2\right) - \lambda \\ \frac{\partial \mathcal{L}}{\partial x_2} &= 2\lambda x_2 + f(\mathbf{x})\left(-2x_1(x_1 x_2 - \frac{3}{2}) - 2(x_2 - \frac{3}{2})\right) \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= x_2^2 - x_1 \end{aligned}$$

Setting these derivatives to zero and solving yields $x_1 \approx 1.358$, $x_2 \approx 1.165$, and $\lambda \approx 0.170$.

Inequality Constraints

- Consider a problem with a single inequality constraint:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g(\mathbf{x}) \leq 0 \end{aligned}$$

- **We know that if the solution lies at the constraint boundary, then the Lagrange condition holds** for some constant μ.

$$\nabla f - \mu \nabla g = \mathbf{0}$$

- When this occurs, the constraint is considered *active*, and the gradient of the objective function is limited exactly as it was with equality constraints. Figure shows an example.

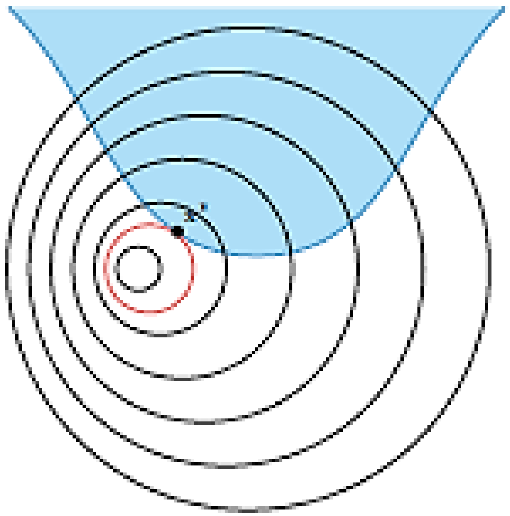


Figure 10.5. An active inequality constraint. The corresponding contour line is shown in red.

Inequality Constraints

- If the solution to the problem does not lie at the constraint boundary, then the constraint is considered *inactive*.
- Solutions of f will simply lie where the gradient of f is zero, as with unconstrained optimization.
- In this case, equation will hold by setting μ to zero.
- Figure shows an example.

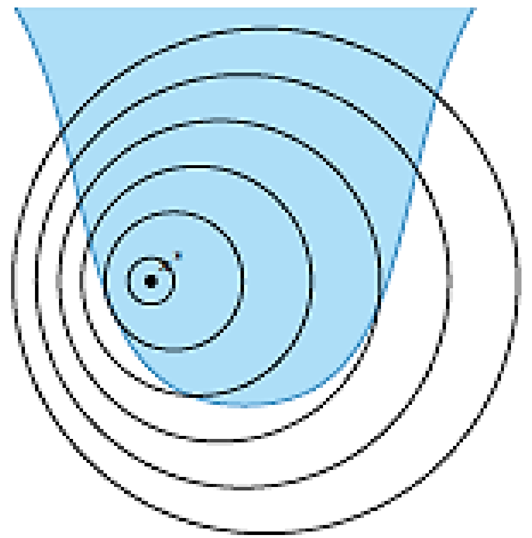


Figure 10.6. An inactive inequality constraint.

Inequality Constraints

- This can be formulated so that the Lagrangian evaluates to infinity outside the feasible set if

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu g(\mathbf{x})$$

then

$$f_{\infty\text{-step}} = \underset{\mu \geq 0}{\text{maximize}}\, \mathcal{L}(\mathbf{x}, \mu)$$

- The new optimization problem becomes

$$\underset{\mathbf{x}}{\text{minimize}}\, \underset{\mu \geq 0}{\text{maximize}}\, \mathcal{L}(\mathbf{x}, \mu)$$

- This reformulation is known as the *primal* problem.

Inequality Constraints

Any primal solution must satisfy the KKT conditions

1. Feasibility:

$$\mathbf{g}(\mathbf{x}^*) \leq 0$$
$$\mathbf{h}(\mathbf{x}^*) = 0$$

2. Dual Feasibility: Penalization is toward feasibility

$$\boldsymbol{\mu} \geq 0$$

3. Complementary Slackness: Either μ_i or $g_i(x^*)$ is zero

$$\boldsymbol{\mu} \odot \mathbf{g} = \mathbf{0}$$

4. Stationarity: Objective function tangent to each active constraint

$$\nabla f(\mathbf{x}^*) - \sum_i \mu_i \nabla g_i(\mathbf{x}^*) - \sum_j \lambda_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

Duality

- The method of Lagrange multipliers can be generalized to define generalized Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_i \mu_i g_i(\mathbf{x}) + \sum_j \lambda_j h_j(\mathbf{x})$$

- As previously mentioned, the primal form is

$$\underset{\mathbf{x}}{\text{minimize}}\ \underset{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}}{\text{maximize}}\ \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

- Reversing the order of operations leads to the dual form

$$\underset{\boldsymbol{\mu} \geq 0, \boldsymbol{\lambda}}{\text{maximize}}\ \underset{\mathbf{x}}{\text{minimize}}\ \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

Duality

- The min-max inequality states that for any function f(**a**,**b**)

$$\underset{\mathbf{a}}{\text{maximize}}\,\underset{\mathbf{b}}{\text{minimize}}\, f(\mathbf{a},\mathbf{b}) \leq \underset{\mathbf{b}}{\text{minimize}}\,\underset{\mathbf{a}}{\text{maximize}}\, f(\mathbf{a},\mathbf{b})$$

- Therefore, the solution to the dual problem is a lower bound to the primal solution
- The dual problem can be used to define the dual function

$$\mathcal{D}(\boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\lambda}) = \underset{\mathbf{x}}{\text{minimize}}\, \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$$

Duality

- By the max-min inequality, the dual solution is a lower bound to the primal solution

$$d^* \leq p^*$$

- The difference between dual and primal solutions d^* and p^* is called the duality gap
- Showing zero-duality gap is a "certificate" of optimality

Duality

Consider the optimization problem:

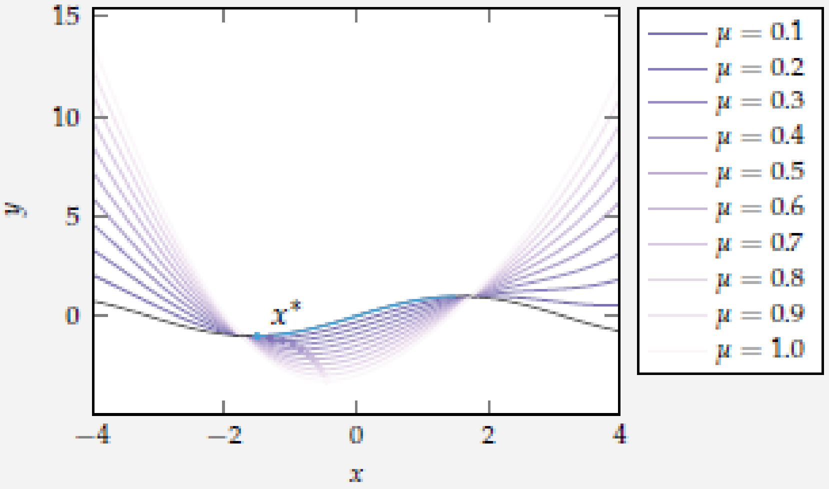
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \sin(x) \\ \text{subject to} \quad & x^2 \leq 3 \end{aligned}$$

The generalized Lagrangian is $\mathcal{L}(x, \mu) = \sin(x) + \mu(x^2 - 3)$, making the primal problem:

$$\underset{x}{\text{minimize}} \; \underset{\mu \geq 0}{\text{maximize}} \; \sin(x) + \mu(x^2 - 3)$$

and the dual problem:

$$\underset{\mu \geq 0}{\text{maximize}} \; \underset{x}{\text{minimize}} \; \sin(x) + \mu(x^2 - 3)$$



The objective function is plotted in black, with the feasible region traced over in blue. The minimum is at $x^* = -1.5$ with $p^* \approx -0.997$. The purple lines are the Lagrangian $\mathcal{L}(x, \mu)$ for $\mu = 0.1, 0.2, \ldots, 1$, each of which has a minimum lower than p^*.

Duality

Consider the problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & x_1 + x_2 + x_1 x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

The Lagrangian is $\mathcal{L}(x_1, x_2, \lambda) = x_1 + x_2 + x_1x_2 + \lambda(x_1^2 + x_2^2 - 1)$.
We apply the method of Lagrange multipliers:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= 1 + x_2 + 2\lambda x_1 = 0 \\ \frac{\partial \mathcal{L}}{\partial x_2} &= 1 + x_1 + 2\lambda x_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= x_1^2 + x_2^2 - 1 = 0 \end{aligned}$$

Solving yields four potential solutions, and thus four critical points:

x_1	x_2	λ
-1	0	$1/2$
0	-1	$1/2$
$\frac{\sqrt{2}+1}{\sqrt{2}+2}$	$\frac{\sqrt{2}+1}{\sqrt{2}+2}$	$\frac{1}{2}\left(-1-\sqrt{2}\right)$
$\frac{\sqrt{2}-1}{\sqrt{2}-2}$	$\frac{\sqrt{2}-1}{\sqrt{2}-2}$	$\frac{1}{2}\left(-1+\sqrt{2}\right)$

The dual function has the form

$$\mathcal{D}(\lambda) = \underset{x_1, x_2}{\text{minimize}}\, x_1 + x_2 + x_1x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

We can substitute in $x_1 = x_2 = x$ and set the derivative with respect to x to zero to obtain $x = -1 - \lambda$. Making the substitution yields

$$\mathcal{D}(\lambda) = -1 - 3\lambda - \lambda^2$$

The dual problem maximize$_\lambda$ $\mathcal{D}(\lambda)$ is maximized at $\lambda = \left(-1 - \sqrt{2}\right)/2$.

Penalty Methods

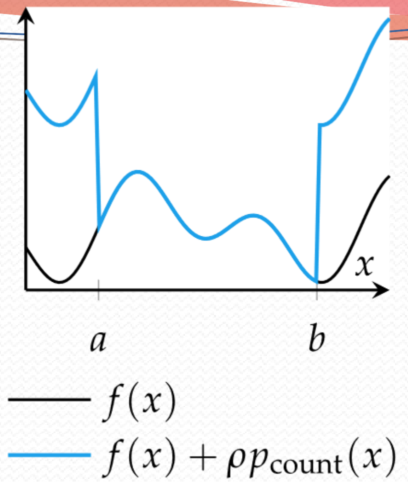
- Penalty methods are a way of reformulating a constrained optimization problem as an unconstrained problem by penalizing the objective function value when constraints are violated
- Simple example

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \le \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{aligned} \quad \longrightarrow \quad \begin{gathered} \underset{\mathbf{x}}{\text{minimize}}\, f(\mathbf{x}) + \rho \cdot p_{\text{count}}(\mathbf{x}) \\ p_{\text{count}}(\mathbf{x}) = \sum_i (g_i(\mathbf{x}) > 0) + \sum_j (h_j(\mathbf{x}) \neq 0) \end{gathered}$$

Penalty Methods

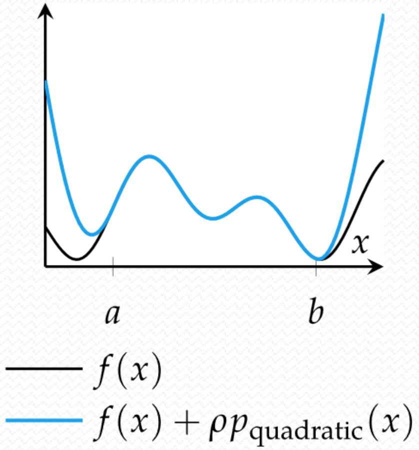
- Count penalty:

$$p_{\text{count}}(\mathbf{x}) = \sum_i (g_i(\mathbf{x}) > 0) + \sum_j (h_j(\mathbf{x}) \neq 0)$$



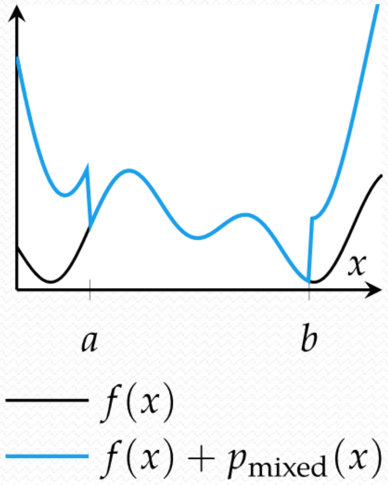
- Quadratic penalty:

$$p_{\text{quadratic}}(\mathbf{x}) = \sum_i \max(g_i(\mathbf{x}), 0)^2 + \sum_j h_j(\mathbf{x})^2$$



- Mixed Penalty:

$$p_{\text{mixed}}(\mathbf{x}) = \rho_1 p_{\text{count}}(\mathbf{x}) + \rho_2 p_{\text{quadratic}}(\mathbf{x})$$



Penalty Methods

Consider the problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x \\ \text{subject to} \quad & x \geq 5 \end{aligned}$$

using a quadratic penalty function.

The unconstrained objective function is

$$f(x) = x + \rho \max(5 - x, 0)^2$$

The minimum of the unconstrained objective function is

$$x^* = 5 - \frac{1}{2\rho}$$

While the minimum of the constrained optimization problem is clearly $x = 5$, the minimum of the penalized optimization problem merely approaches $x = 5$, requiring an infinite penalty to achieve feasibility.

Augmented Lagrange Method

- The *augmented Lagrange method*11 is an adaptation of the penalty method specifically for equality constraints. Unlike the penalty method,where ρ must sometimes approach infinity before a feasible solution is found, the augmented Lagrange method will work with smaller values of ρ. It uses both a quadratic and a linear penalty for each constraint

$$p_{\text{Lagrange}}(\mathbf{x}) = \frac{1}{2}\rho \sum_i (h_i(\mathbf{x}))^2 - \sum_i \lambda_i h_i(\mathbf{x})$$

- λ converges towards the Lagrange multiplier

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} - \rho \mathbf{h}(\mathbf{x})$$

Interior Point Methods

- Also called barrier methods, interior point methods ensure that each step is feasible
- This allows premature termination to return a nearly optimal, feasible point
- Barrier functions are implemented similar to penalties but must meet the following conditions

1. Continuous
2. Non-negative
3. Approach infinity as **x** approaches any constraint boundary

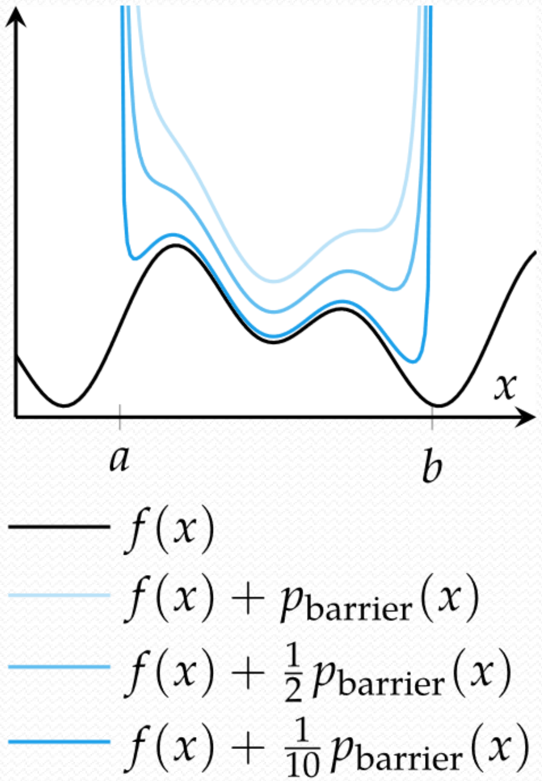
Interior Point Methods

- **Inverse Barrier:** $$p_{\text{barrier}}(\mathbf{x}) = -\sum_i \frac{1}{g_i(\mathbf{x})}$$

- **Log Barrier:** $$p_{\text{barrier}}(\mathbf{x}) = -\sum_i \begin{cases} \log(-g_i(\mathbf{x})) & \text{if } g_i(\mathbf{x}) \geq -1 \\ 0 & \text{otherwise} \end{cases}$$

- New optimization problem

$$\underset{\mathbf{x}}{\text{minimize}}\, f(\mathbf{x}) + \frac{1}{\rho} p_{\text{barrier}}(\mathbf{x})$$



Summary

- Constraints are requirements on the design points that a solution must satisfy
- Some constraints can be transformed or substituted into the problem to result in an unconstrained optimization problem
- Analytical methods using Lagrange multipliers yield the generalized Lagrangian and the necessary conditions for optimality under constraints
- A constrained optimization problem has a dual problem formulation that is easier to solve and whose solution is a lower bound of the solution to the original problem

Summary

- Penalty methods penalize infeasible solutions and often provide gradient information to the optimizer to guide infeasible points toward feasibility
- Interior point methods maintain feasibility but use barrier functions to avoid leaving the feasible set