Constrained Optimization: Introduction and Karush-Kuhn-Tucker Conditions

AE 6310: Optimization for the Design of Engineered Systems

Spring 2017

Dr. Glenn Lightsey

Lecture Notes Developed By Dr. Brian German

Georgia Institute of Technology

Daniel Guggenheim School of Aerospace Engineering

The Unconstrained Optimization Problem

$$\text{Minimize} \quad f(\boldsymbol{x})$$

- If we wish to maximize $f(\boldsymbol{x})$, then we can instead minimize $-f(\boldsymbol{x})$
- The function $f(\boldsymbol{x})$ is of unknown form; we have little or no knowledge about its shape
- $f(\boldsymbol{x})$ may be expensive to evaluate; we should "query" it as little as possible to reduce the number of "function calls"

The Constrained Optimization Problem

Minimize $f(\boldsymbol{x})$

Subject to:

$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$	Inequality constraints
$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$	Equality constraints
$\boldsymbol{x}_L \leq \boldsymbol{x} \leq \boldsymbol{x}_U$	Side constraints

This formulation is called a *standard form* for constrained optimization. Any particular problem could be arranged into this form.

Active and Inactive Constraints

An inequality constraint is said to be *active* at a potential solution $\boldsymbol{x}^*$ if it satisfies,

$$g(\boldsymbol{x}^*) = 0.$$

An inequality constraint is said to be *inactive* or *passive* at a potential solution $\boldsymbol{x}^*$ if it satisfies,

$$g(\boldsymbol{x}^*) < 0.$$

Equality constraints are always active.

Feasible and Infeasible Point

A point $\boldsymbol{x}^*$ is said to be *feasible* if it satisfies all of the constraints.

A point $\boldsymbol{x}^*$ is said to be *infeasible* if one or more of the constraints is not satisfied.

Example Constrained Optimization Problem

Minimize: $f(x_1, x_2) = \frac{3}{2}x_1^2 + \frac{1}{2}x_1x_2 + 2x_2^2 - \frac{1}{2}x_1 + \frac{1}{2}x_2 - 2$

Subject to:

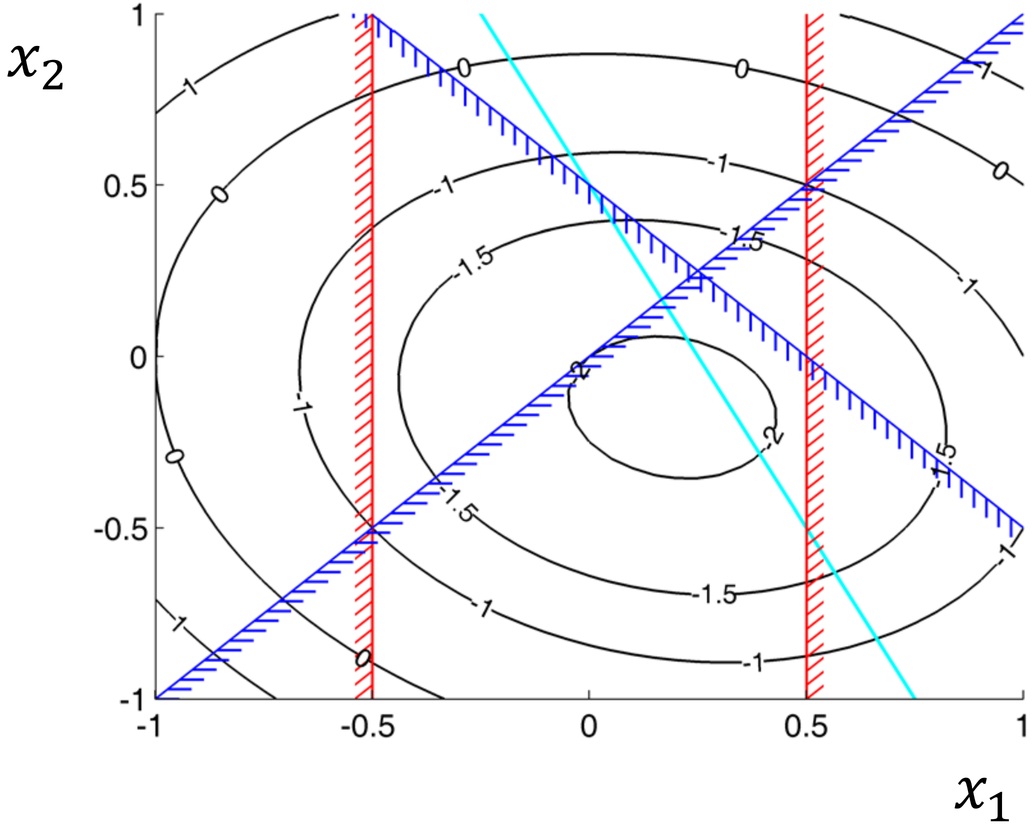
$$g_1(x_1, x_2) = x_1 - x_2 \leq 0$$

$$g_2(x_1, x_2) = -x_1 - x_2 + \frac{1}{2} \leq 0$$

$$h_1(x_1, x_2) = 2x_1 + x_2 - \frac{1}{2} = 0$$

$$-0.5 \leq x_1 \leq 0.5$$

Example Constrained Optimization Problem



Where is the minimum?

The Lagrangian Function

Analysis of constrained optimization is based upon the *Lagrangian function:*

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}) + \sum_{k=1}^{l} \lambda_{m+k} h_k(\boldsymbol{x})$$

The Lagrangian "modifies" the objective function to account for the *m* inequality constraints (including side constraints) and *l* equality constraints.

Lagrangian Multipliers

The λ_j and λ_{m+k} factors that appear in the Lagrangian are called *Lagrange multipliers.*

We will soon see more about what they mean.

First-Order Necessary Conditions

Recall that for unconstrained optimization, the first-order necessary condition for the optimum was as follows:

Let $f(\boldsymbol{x})$ be a function that is once differentiable at $\boldsymbol{x}^*$:

- If $\boldsymbol{x}^*$ is a local solution to the unconstrained minimization problem, then $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

For constrained optimization, the necessary condition must account for the constraints. The corresponding conditions are called the *Karush-Kuhn-Tucker Conditions.*

The Karush-Kuhn-Tucker (KKT) Conditions

1. $\boldsymbol{x}^*$ is feasible

2. $\lambda_j g_j(\boldsymbol{x}^*) = 0, \quad j = 1, \dots, m$ (Complementary Slackness)

3. $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(\boldsymbol{x}^*) + \sum_{k=1}^{l} \lambda_{m+k} \nabla_x h_k(\boldsymbol{x}^*) = \boldsymbol{0}$

where $\lambda_j \geq 0, \quad j = 1, \dots, m$

and λ_{m+k} are unrestricted in sign, $\quad k = 1, \dots, l$

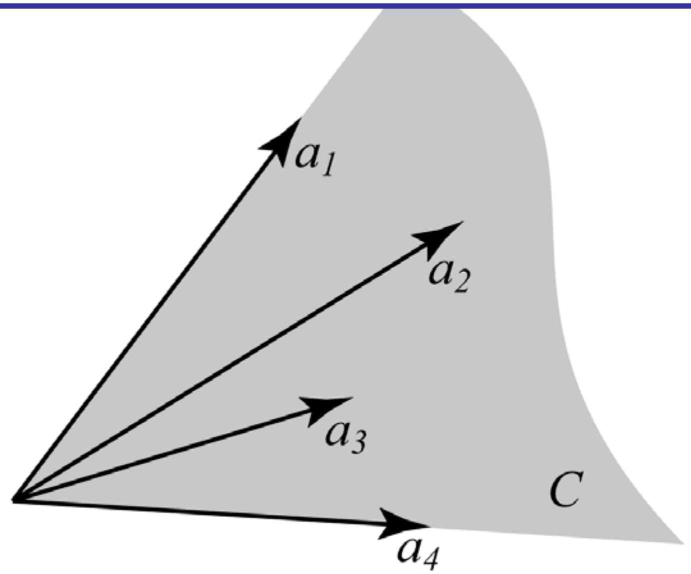
Implications of $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = 0$

$$\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(\boldsymbol{x}^*)$$

This is a vector, pointing in the "steepest ascent" direction of the objective function

This is a linear combination of vectors, with non-negative "weights". The set of all such vectors is called a **convex cone.**

Geometrically, a convex cone is the linear subspace spanned by the vectors. In other words, it is the convex hull of the vectors.



Implications of $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = 0$

$$\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(\boldsymbol{x}^*)$$

- For a particular setting of the Lagrange multipliers, the second term is some vector in the convex cone.
- So we can interpret this equation as: "In order for $\boldsymbol{x}^*$ to be a constrained minimum, the negative of the gradient of the objective function must be equal to some vector *in* the convex cone defined by the gradients of the inequality constraints.
- In other words, the negative of the gradient of the objective function must be *in* the convex cone defined by the gradients of the inequality constraints.
- Equivalently, we can say that the gradient of the objective function must be in the convex cone defined by the negative gradients of the inequality constraints.

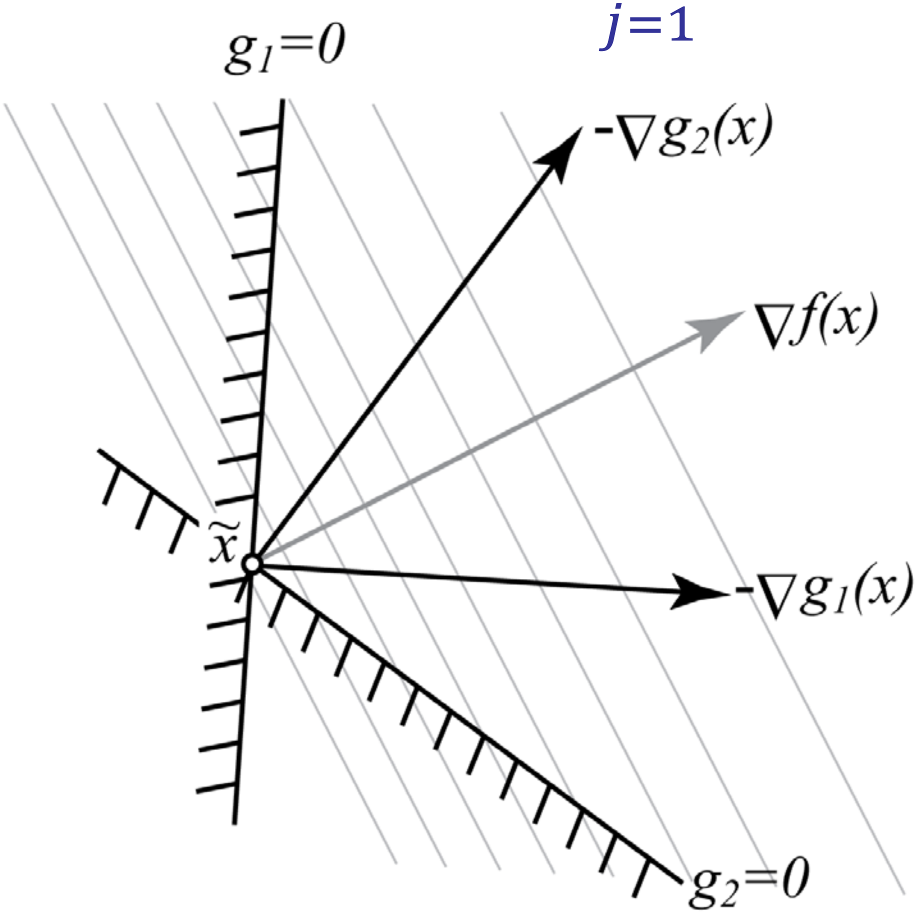
Implications of $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = 0$

$$\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(\boldsymbol{x}^*)$$

So, this equation is setting a requirement on the direction of the objective function's gradient relative to the direction of the inequality constraints' gradients.

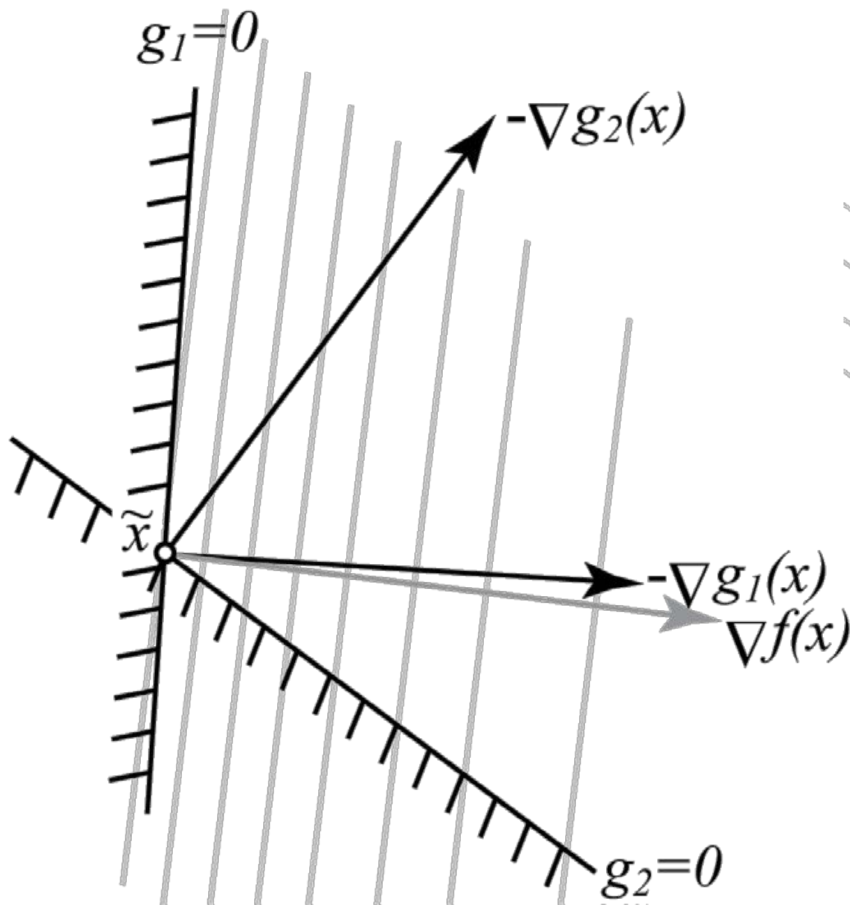
Implications of $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = 0$

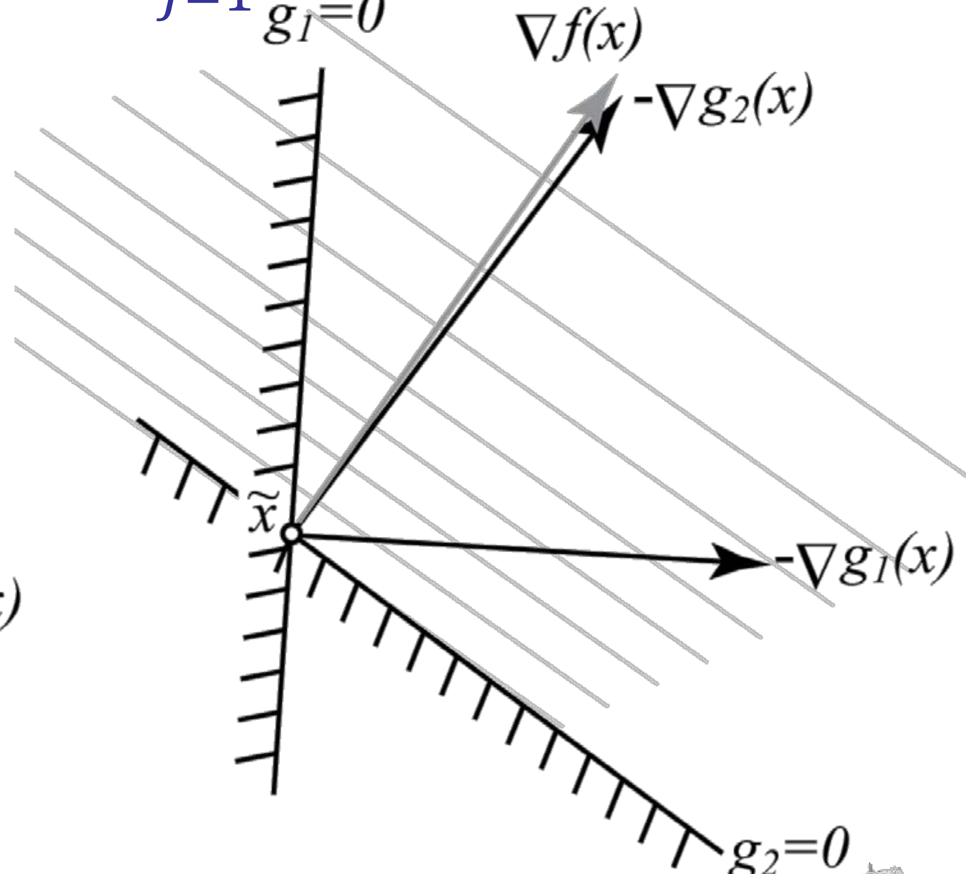
$$\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(\boldsymbol{x}^*)$$



What if $\nabla_x f(\boldsymbol{x}^*)$ is outside the convex cone?

$$\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla_x f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \nabla_x g_j(\boldsymbol{x}^*)$$

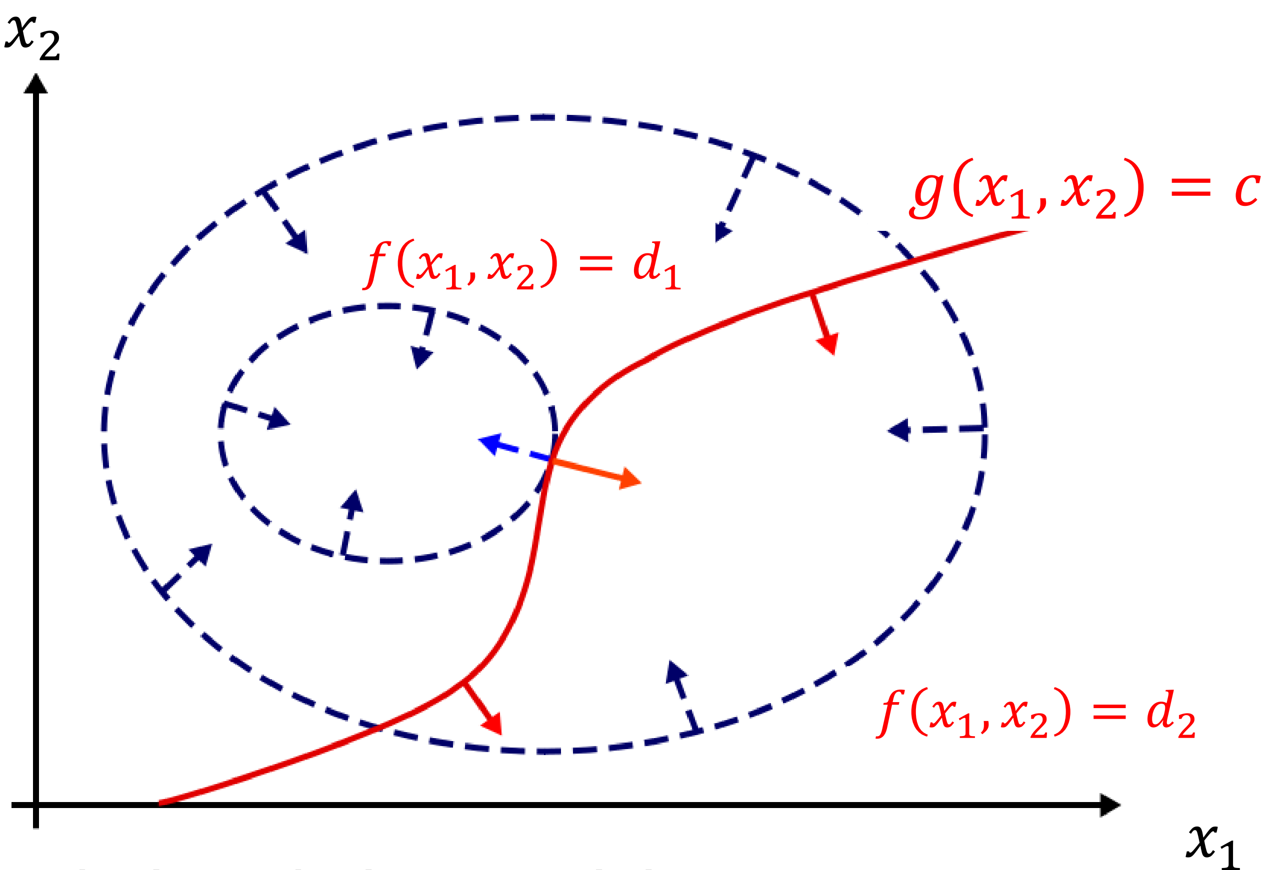




Implications of $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}) = 0$

$$\nabla \mathcal{L}_x(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \nabla f(\boldsymbol{x}^*) + \lambda \nabla g(\boldsymbol{x}^*) = \mathbf{0}$$

If there is **only one** active inequality constraint, the convex cone collapses to a line, so the objective function's gradient must point in exactly the opposite direction as the inequality constraint's gradient



Adapted from http://en.wikipedia.org/wiki/File:LagrangeMultipliers2D.svg

Example Application of the KKT Conditions

In some cases, the KKT conditions can be applied directly to solve a constrained optimization problem. Consider the following example:

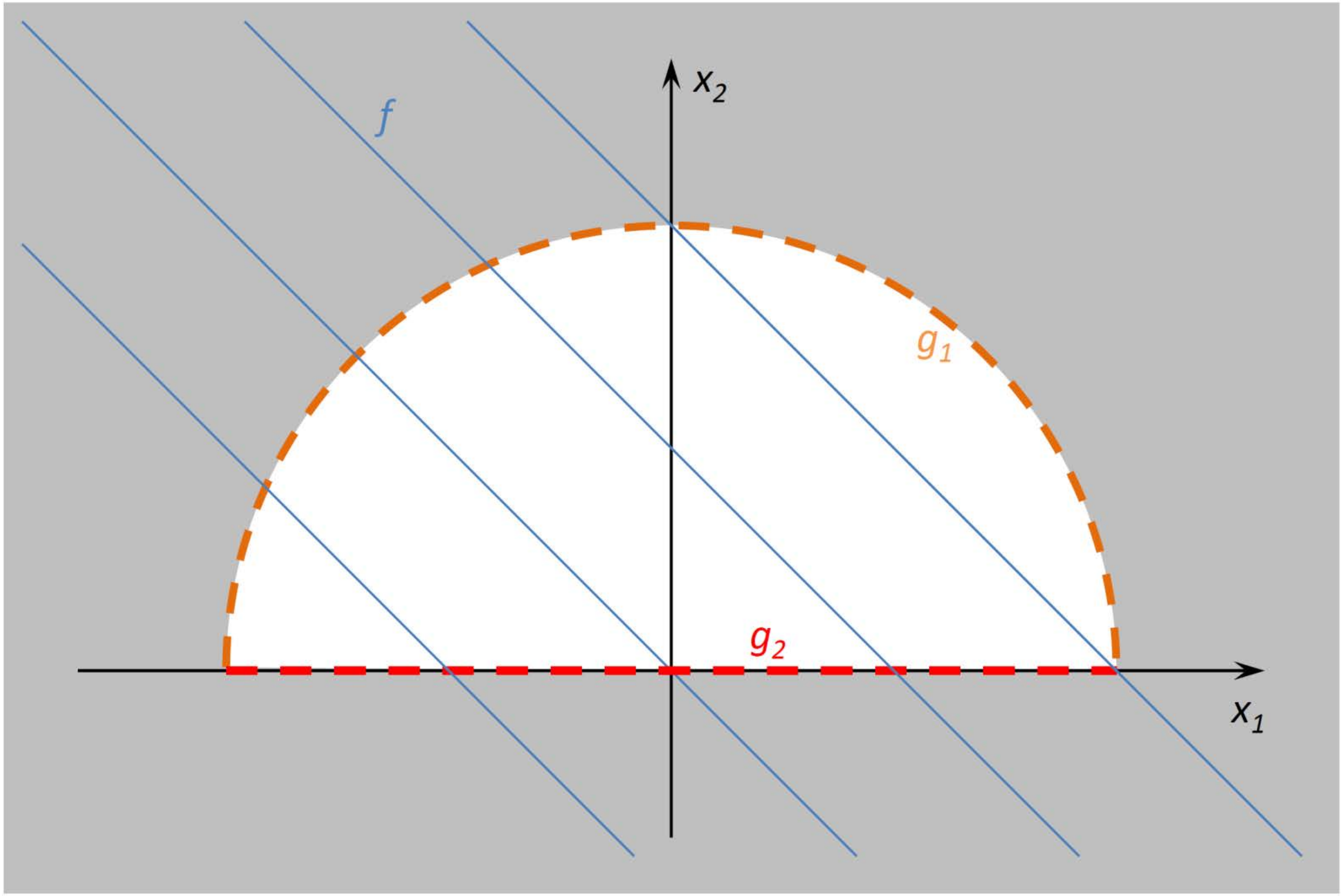
Minimize: $f(x_1, x_2) = x_1 + x_2$

Subject to:

$$g_1(x_1, x_2) = x_1^2 + x_2^2 - 1 \leq 0$$

$$g_2(x_1, x_2) = -x_2 \leq 0$$

Example Application of the KKT Conditions



Example Application of the KKT Conditions

First, let's form the Lagrangian:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}) + \sum_{k=1}^{l} \lambda_{m+k} h_k(\boldsymbol{x})$$

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2) = (x_1 + x_2) + \lambda_1(x_1^2 + x_2^2 - 1) + \lambda_2(-x_2)$$

Example Application of the KKT Conditions

First, let's apply KKT condition 2:

$$\lambda_1(x_1^2 + x_2^2 - 1) = 0$$

$$\lambda_2(-x_2) = 0$$

Next, let's apply KKT condition 3:

$$\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2) = (x_1 + x_2) + \lambda_1(x_1^2 + x_2^2 - 1) + \lambda_2(-x_2)$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = 1 + 2\lambda_1 x_1 = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 1 + 2\lambda_1 x_2 - \lambda_2 = 0$$

Example Application of the KKT Conditions

We can now solve these four equations for x_1, x_2, λ_1 and λ_2 as,

$$x_1 = \pm 1$$

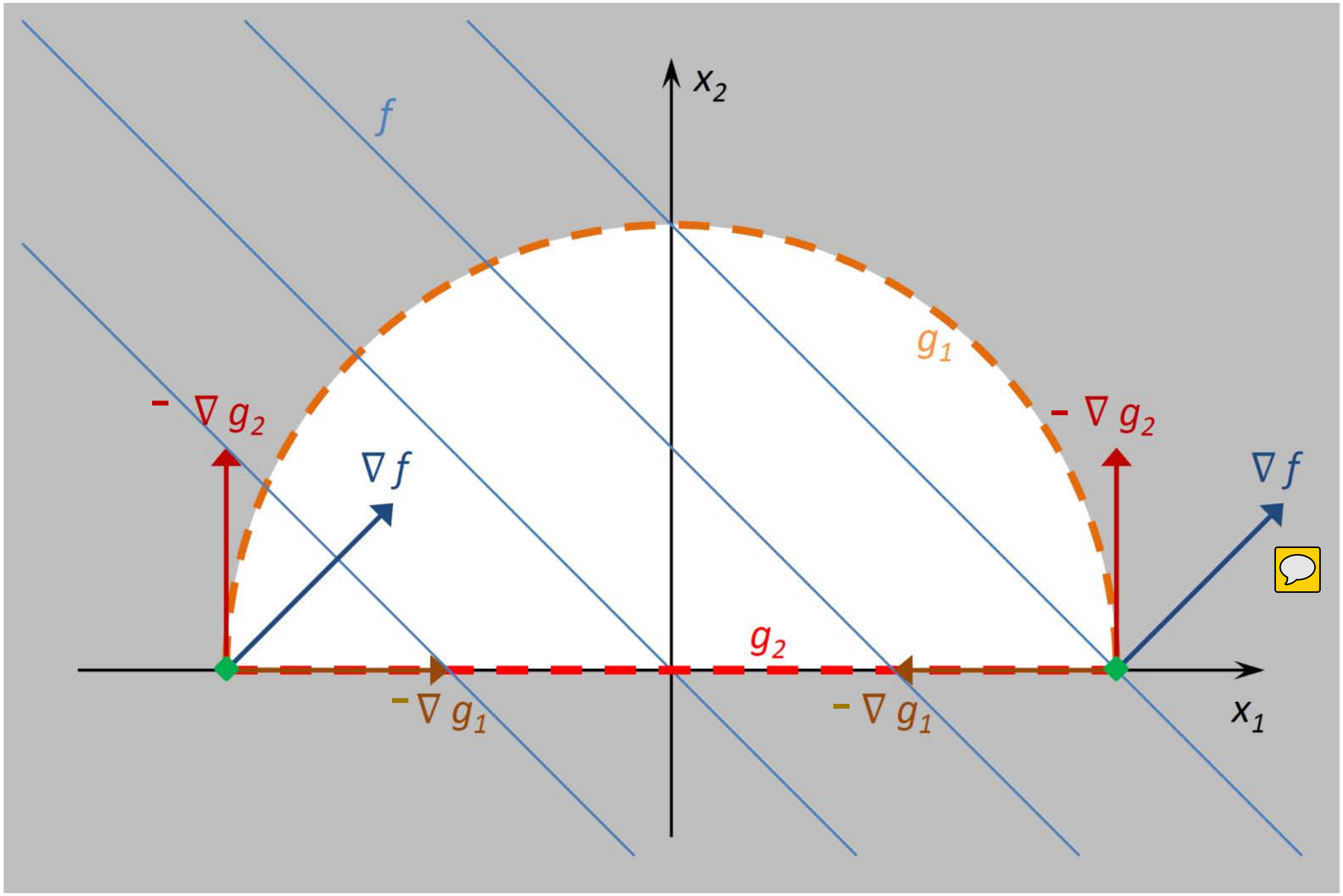
$$x_2 = 0$$

$$\lambda_1 = -\frac{1}{2x_1}$$

$$\lambda_2 = 1$$

There are two points that satisfy KKT conditions 2 and 3. Which, if either, could be the constrained minimum?

Example Application of the KKT Conditions



The solution corresponds to the *positive* λ_1 which occurs at $x_1 = -1, x_2 = 0$

Example Application of the KKT Conditions

- A note of caution: Even if you can "solve" a constrained optimization problem using the KKT conditions to find a point $\boldsymbol{x}^*$, we are not guaranteed that the point is a local optimum.

- The reason is that the KKT conditions are only "necessary" conditions for a constrained local optimum. We would need to meet the "sufficient" conditions to guarantee a local optimum. (More on the sufficient conditions later…)

Do the KKT Conditions Always Apply?

- No.
- The KKT Conditions apply only if the constraints satisfy some regularity conditions called "**constraint qualifications**."
- A commonly applied condition is that the gradients of the active constraints must be linearly independent at the point at which the KKT conditions are being checked.
- Other less restrictive conditions also exist

Do the KKT Conditions Always Apply?

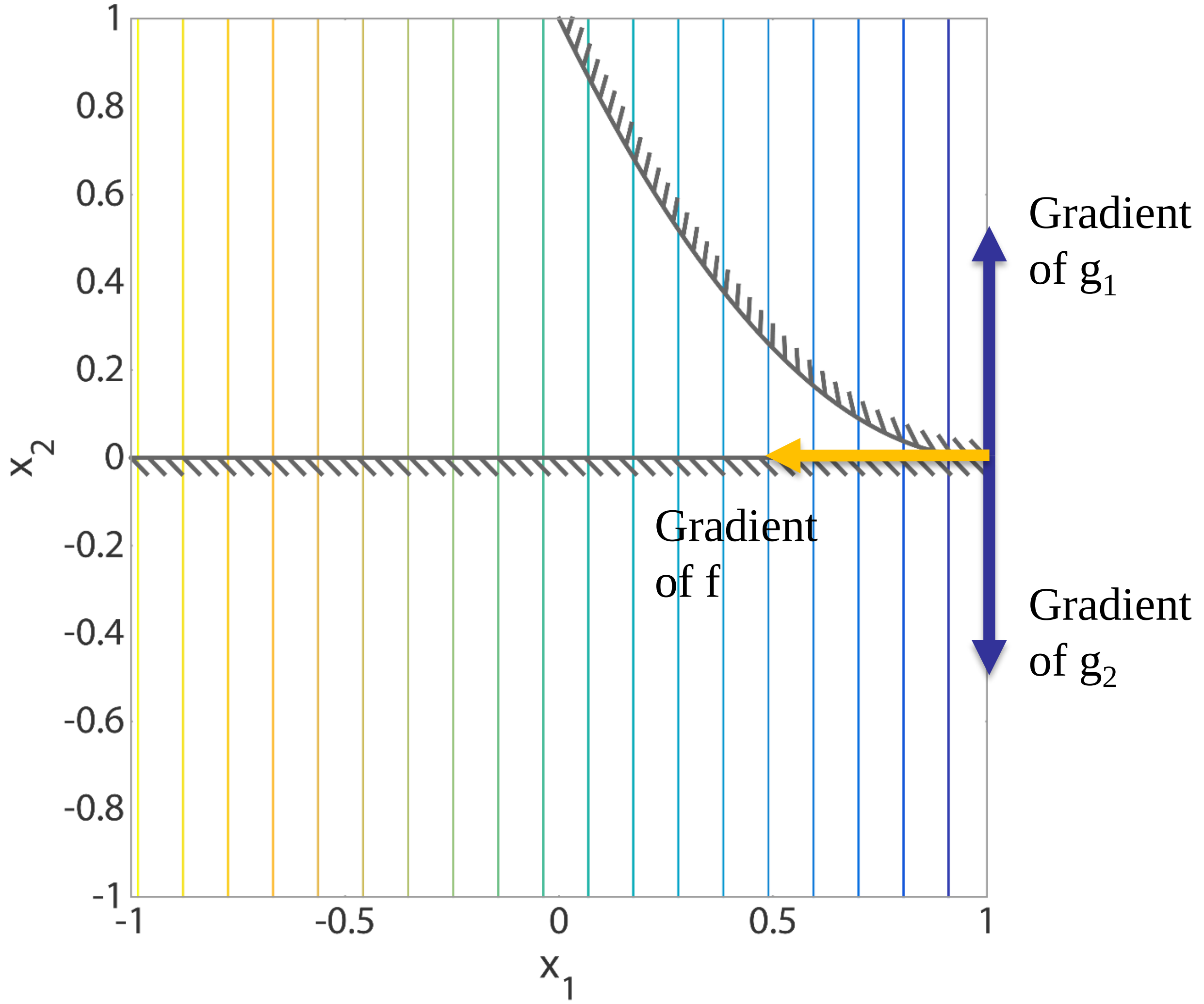
Consider the following problem

Minimize $f = -x_1$

s.t.

$$x_2 \geq 0$$
$$x_2 + (x_1 - 1)^3 \leq 0$$

Do the KKT Conditions Always Apply?



Second-Order Necessary/Sufficient Conditions

Read Nocedal and Wright, Section 12.5. (It wouldn't hurt to read 12.3 and 12.4 as well.)

Approaches to Constrained Optimization

There are two general types of approaches for solving constrained optimization problems:

- **Indirect methods** modify the objective function with a *penalty function* to account for the influence of the constraints. The problems are then solved with unconstrained optimization techniques.

- **Direct methods** enforce the constraints explicitly to achieve a solution. These methods therefore depart considerably from unconstrained optimization techniques.