Plan for today

- Nonlinear Optimization
  - Unconstrained
    - Verifying convexity
    - Minimizing Convex Functions
  - Constrained
    - Equality constraints: Lagrangian Method
    - Inequality constraints: KKT Conditions Method

NONLINEAR OPTIMIZATION

… where we see nonlinear optimization models and define convex functions

Nonlinear Optimization = Non Linear Programming (NLP)

Linear vs Non-Linear functions

- A function $f(\boldsymbol{x})$ is **linear** if it satisfies proportionality and additivity

$$f(x) = 7x - 4$$
$$f(\boldsymbol{x}) = 3x_1 + 7x_2 - 15x_3$$

- A function $f(\boldsymbol{x})$ is **non-linear** if it violates proportionality or additivity

$$f(x) = x^2 - x + 4$$
$$f(\boldsymbol{x}) = x_1^2 - 16x_1^3 + x_1x_2$$

NLP: UNCONSTRAINED

Basic Concepts in NLP

Local vs Global Minima

- **Local Minimum**:
  - A point is a local minimum
    if there are no better (feasible) solutions "nearby"
  - Formally: A point $\boldsymbol{x}^*$ is a local minimum if there exists some $\epsilon > 0$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon$

- **Global Minimum**:
  - A point is a global minimum
    if there are no better (feasible) solutions anywhere
  - Formally: A point $\boldsymbol{x}^*$ is a global minimum if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$

How to find stationary points for multivar functions?

- Suppose that our mathematical program is of the form

$$\min f(\boldsymbol{x})$$
$$\boldsymbol{x} \text{ unrestricted}$$

where $f(\boldsymbol{x})$ is a multi var differentiable function

- From calculus, we can find stationary points by solving for x from

$$\boldsymbol{\nabla f}(\boldsymbol{x}) = 0$$

where the gradient of f is defined as $\boldsymbol{\nabla f}(\boldsymbol{x}) := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)$

- Checking if a stationary point is a local min/max is much harder for multivar functions

Local vs Global Minima/Maxima

- **Local Minimum**:
  - A point is a local minimum
    if there are no better (feasible) solutions "nearby"
  - Formally: A point $\boldsymbol{x}^*$ is a local minimum if there exists some $\epsilon > 0$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$ such that $||\boldsymbol{x} - \boldsymbol{x}^*|| < \epsilon$

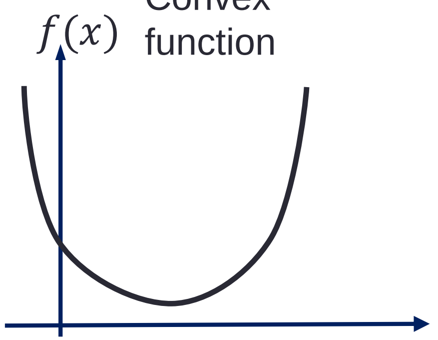
- **Global Minimum**:
  - A point is a global minimum
    if there are no better (feasible) solutions anywhere
  - Formally: A point $\boldsymbol{x}^*$ is a global minimum if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$

Question: Suppose we know that $\boldsymbol{x}^*$ is a stationary point.
Can we tell whether it is a global minimum?

Convexity helps!

- If $f(\boldsymbol{x})$ is a *convex function*, then any stationary point is a global minimum
- If $f(\boldsymbol{x})$ is a *concave function*, then any stationary point is a global maximum

Convex Functions



A function $f(\boldsymbol{x})$ is a

1. **convex function** if

   for each pair of points $\boldsymbol{a}$ and $\boldsymbol{b}$ and every $0 < \lambda < 1$:

$$f(\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b}) \leq \lambda f(\boldsymbol{a}) + (1-\lambda) f(\boldsymbol{b})$$

2. **strict convex function** if

   for each pair of points $\boldsymbol{a}$ and $\boldsymbol{b}$ and every $0 < \lambda < 1$:

$$f(\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b}) < \lambda f(\boldsymbol{a}) + (1-\lambda) f(\boldsymbol{b})$$

3. **concave function** if

   for each pair of points $\boldsymbol{a}$ and $\boldsymbol{b}$ and every $0 < \lambda < 1$:

$$f(\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b}) \geq \lambda f(\boldsymbol{a}) + (1-\lambda) f(\boldsymbol{b})$$

4. **strict concave function** if

   for each pair of points $\boldsymbol{a}$ and $\boldsymbol{b}$ and every $0 < \lambda < 1$:

$$f(\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b}) > \lambda f(\boldsymbol{a}) + (1-\lambda) f(\boldsymbol{b})$$

Local vs Global Minima/Maxima

- **Local Minimum**:
  - A point is a local minimum
    if there are no better (feasible) solutions "nearby"
  - Formally: A point $\boldsymbol{x}^*$ is a local minimum if there exists some $\epsilon > 0$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$ such that $||\boldsymbol{x} - \boldsymbol{x}^*|| < \epsilon$

- **Global Minimum**:
  - A point is a global minimum
    if there are no better (feasible) solutions anywhere
  - Formally: A point $\boldsymbol{x}^*$ is a global minimum if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$

Question: Suppose we know that $\boldsymbol{x}^*$ is a stationary point.
Can we tell whether it is a global minimum?

Convexity helps!

- If $f(\boldsymbol{x})$ is a *convex function*, then any stationary point is a global minimum
- If $f(\boldsymbol{x})$ is a *concave function*, then any stationary point is a global maximum

HOW TO VERIFY CONVEXITY?

- By definition (often hard)
- Derivative tests

How to verify convexity for multi var functions?

- For a single variable twice differentiable function $f(x)$
  If $f''(x) \geq 0$ for all x, then $f(x)$ is a convex function
  If $f''(x) > 0$ for all x, then $f(x)$ is a strictly convex function
  If $f''(x) \leq 0$ for all x, then $f(x)$ is a concave function
  If $f''(x) < 0$ for all x, then $f(x)$ is a strictly concave function

- A multi variable twice-differentiable function $f(\boldsymbol{x})$ is convex if all eigenvalues of the **Hessian matrix** $\boldsymbol{\nabla^2 f(x)}$ are non-negative
  - The i-th row, j-th col entry of the *Hessian matrix* is obtained by
$$\left[\boldsymbol{\nabla^2 f(x)}\right]_{i,j} := \frac{\partial^2 f}{\partial x_i \partial x_j}$$
  - Recall: Eigenvalues of a matrix M are the roots of the equation
$$\det(M - yI) = 0$$

PUTTING THE CONCEPTS TOGETHER

Unconstrained Minimization

Minimizing Single Variable Twice-Differentiable Functions

$$\min f(x)$$
$$x \text{ unrestricted}$$

1. Find a stationary point x^*
   … by solving $f'(x^*) = 0$

2. Verify that $f(x)$ is convex
   … by verifying if $f''(x) > 0$ for all x

Minimizing Single Variable Twice-Differentiable Functions

$$\min f(x) = x^2$$
$$x \text{ unrestricted}$$

1. Find a stationary point x^*
   … by solving $f'(x^*) = 0$

$$f'(x) = 2x$$
$$x^* = 0$$

2. Verify that $f(x)$ is convex
   … by verifying if $f''(x) > 0$ for all x

… Have done this already for this example

Minimizing Multi Variable Twice-Differentiable Functions

$$\min f(\boldsymbol{x})$$
$$\boldsymbol{x} = (x_1, x_2, \dots, x_n) \text{ unrestricted}$$

1. Find a point x^* where the gradient is zero
   … by solving $\boldsymbol{\nabla f}(\boldsymbol{x}) = 0$ where,
   the gradient of f is defined as $\boldsymbol{\nabla f}(\boldsymbol{x}) := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)$

2. Verify that $f(\boldsymbol{x})$ is convex
   … by verifying if all eigenvalues of the
   Hessian matrix $\boldsymbol{\nabla^2 f}(\boldsymbol{x})$ are non-negative

Minimizing Multi Variable Twice-Differentiable Functions

$$\min f(x_1, x_2) = x_1^2 + x_2^2$$
$$\boldsymbol{x} = (x_1, x_2) \text{ unrestricted}$$

1. Find a point x^* where the gradient is zero

   … by solving $\boldsymbol{\nabla f}(\boldsymbol{x}) = 0$ where,

   the gradient of f is defined as $\boldsymbol{\nabla f}(\boldsymbol{x}) \coloneqq \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$

$$\boldsymbol{\nabla f}(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1} = 2x_1, \frac{\partial f}{\partial x_2} = 2x_2\right)$$

$$\boldsymbol{x}^* = (x_1^* = 0 \quad , x_2^* = 0 \ )$$

2. Verify that $f(\boldsymbol{x})$ is convex

   … by verifying if all <u>eigenvalues</u> of the **Hessian matrix** $\boldsymbol{\nabla^2 f}(\boldsymbol{x})$ are non-negative

… Have done this already for this example

Minimizing Single Variable Twice-Differentiable Functions

$$\begin{aligned} &\min f(x) \\ &x \text{ unrestricted} \end{aligned}$$

1. Find a stationary point x^*

   … by solving $f'(x^*) = 0$

2. Verify that $f(x)$ is convex

   … by verifying if $f''(x) > 0$ for all x

If $f(x)$ is <u>not</u> convex, then
1. Find all stationary points and identify local minimizers
2. Evaluate the function at all local minimizers to identify the global minimizer

Minimizing Multi Variable Twice-Differentiable Functions

$$\min f(\boldsymbol{x})$$
$$\boldsymbol{x} = (x_1, x_2, \dots, x_n) \text{ unrestricted}$$

1. Find a point x^* where the gradient is zero

   … by solving $\nabla \boldsymbol{f}(\boldsymbol{x}) = 0$ where,

   the gradient of f is defined as $\nabla \boldsymbol{f}(\boldsymbol{x}) \coloneqq \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)$

2. Verify that $f(\boldsymbol{x})$ is convex

   … by verifying if all <u>eigenvalues</u> of the **Hessian matrix** $\nabla^2 \boldsymbol{f}(\boldsymbol{x})$ are non-negative

If $f(\boldsymbol{x})$ is <u>not</u> convex, then problem is very difficult

Machine Learning Applications!

Maximizing vs Minimizing
Convex vs Concave

- Maximizing $f(\boldsymbol{x})$ is equivalent to minimizing $-f(\boldsymbol{x})$
- Also note that $f(\boldsymbol{x})$ is concave
  if and only if
  $-f(\boldsymbol{x})$ is convex
- So, concave maximization is equivalent to convex minimization

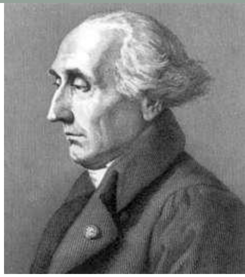
NLP: CONSTRAINED

… where we see optimality conditions for multivariate constrained nonlinear optimization problems

NLP: CONSTRAINED

With only **equality** constraints

- The Lagrangian Method



Constrained Optimization with Equality Constraints

- Suppose we have an optimization problem of the following type:

$$\max f(\boldsymbol{x})$$
$$g_i(\boldsymbol{x}) = b_i \text{ for i} = 1, \dots, m$$

where $f(\boldsymbol{x})$ and any of the $g_i(\boldsymbol{x})$ may be non-linear and $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$

- Given this problem, we can write a related unconstrained optimization problem:

$$\max_{x,\lambda} h(\boldsymbol{x}, \boldsymbol{\lambda})$$
$$h(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i (g_i(\boldsymbol{x}) - b_i)$$

where we have added m new variables $\lambda_1, \lambda_2, \dots, \lambda_m$ called **Lagrange multipliers**

- The objective function $h(\boldsymbol{x}, \boldsymbol{\lambda})$ is called the **Lagrange function**

Constrained Optimization with Equality Constraints

$$\max f(\boldsymbol{x})$$
$$g_i(\boldsymbol{x}) = b_i \text{ for i} = 1, \dots, m$$

$$\max_{x,\lambda} h(\boldsymbol{x}, \boldsymbol{\lambda})$$
$$h(\boldsymbol{x}, \boldsymbol{\lambda}) \coloneqq f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i (g_i(\boldsymbol{x}) - b_i)$$

- We can find the stationary points of $h(\boldsymbol{x}, \boldsymbol{\lambda})$ in the same way that we found the stationary points of unconstrained multi-var NLP
  - We now have $n + m$ decision variables in $h(\boldsymbol{x}, \boldsymbol{\lambda})$
  - So take the gradient of h, set it to zero, and solve for x and λ
  - Suppose $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = (x_1^*, x_2^*, \dots, x_n^*, \lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$ is a stationary point
  - Recall that $\frac{\partial h(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial \lambda_i} = g_i(\boldsymbol{x}) - b_i = 0$ at any stationary point of $h(\boldsymbol{x}, \boldsymbol{\lambda})$
  - So every stationary point corresponds to a feasible solution to our constrained optimization problem

- The stationary points of the Lagrangian are known as the **critical points** of the constrained optimization problem

- Note: If $\boldsymbol{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ is a constrained optimum, then there must be some $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$ such that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a stationary point of the Lagrangian
  - I.e., we know that one of the stationary points should be an optimum for the constrained problem

- After finding the stationary points of the Lagrangian, pick the maximizer among them
  - Caution: Remember to verify if the picked point is a maximizer/minimizer

Constrained Optimization with Equality Constraints

$$\max f(\boldsymbol{x})$$
$$g_i(\boldsymbol{x}) = b_i \text{ for i} = 1, \dots, m$$

$$\max_{x,\lambda} h(\boldsymbol{x}, \boldsymbol{\lambda})$$
$$h(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i (g_i(\boldsymbol{x}) - b_i)$$

- Consider the following constrained NLP:

$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 = 1$$

- The Lagrangian function is $h(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1 + x_2 - \lambda_1(x_1^2 + x_2^2 - 1)$

- Find stationary points by setting $\nabla h(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$:
  - $\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_1} = 1 - 2\lambda_1 x_1 = 0$
  - $\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_2} = 1 - 2\lambda_1 x_2 = 0$
  - $\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \lambda_1} = -(x_1^2 + x_2^2 - 1) = 0$

- We get a system of nonlinear equations, which may be difficult to solve

Constrained Optimization with Equality Constraints

$$\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_1} = 1 - 2\lambda_1 x_1 = 0$$
$$\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_2} = 1 - 2\lambda_1 x_2 = 0$$
$$\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \lambda_1} = -(x_1^2 + x_2^2 - 1) = 0$$

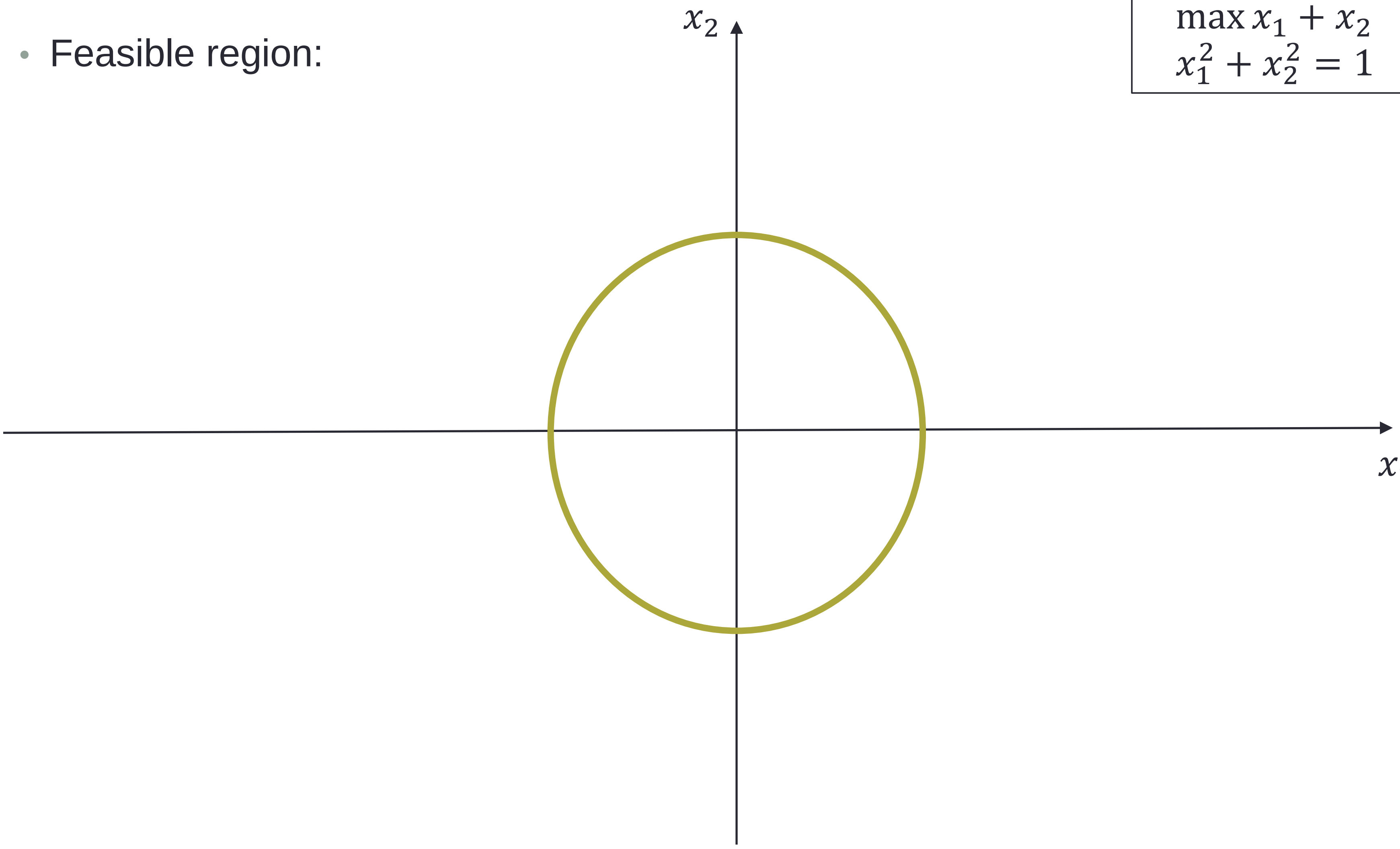
- From the first two equations, we have:
  - $\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_1} = 1 - 2\lambda_1 x_1 = 0$, implies $x_1^* = \frac{1}{2\lambda_1^*}$
  - $\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial x_2} = 1 - 2\lambda_1 x_2 = 0$, implies $x_2^* = \frac{1}{2\lambda_1^*}$
- Substituting into the third equation, we have:
  - $\frac{\partial h(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \lambda_1} = -\left({x_1^*}^2 + {x_2^*}^2 - 1\right) = -\left(\frac{1}{(2\lambda_1^*)^2} + \frac{1}{(2\lambda_1^*)^2} - 1\right) = 0$, implies $\lambda_1^* = \pm\frac{1}{\sqrt{2}}$
- Taking each of the two possible values for λ_1^*, we find two stationary points
  - $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, i.e., $(x_1^*, x_2^*) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$
  - $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, i.e., $(x_1^*, x_2^*) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$
- The stationary point with the best objective value is $(x_1^*, x_2^*) =$

  Is this a maximizer or a minimizer?

Constrained Optimization with Equality Constraints

- Feasible region:

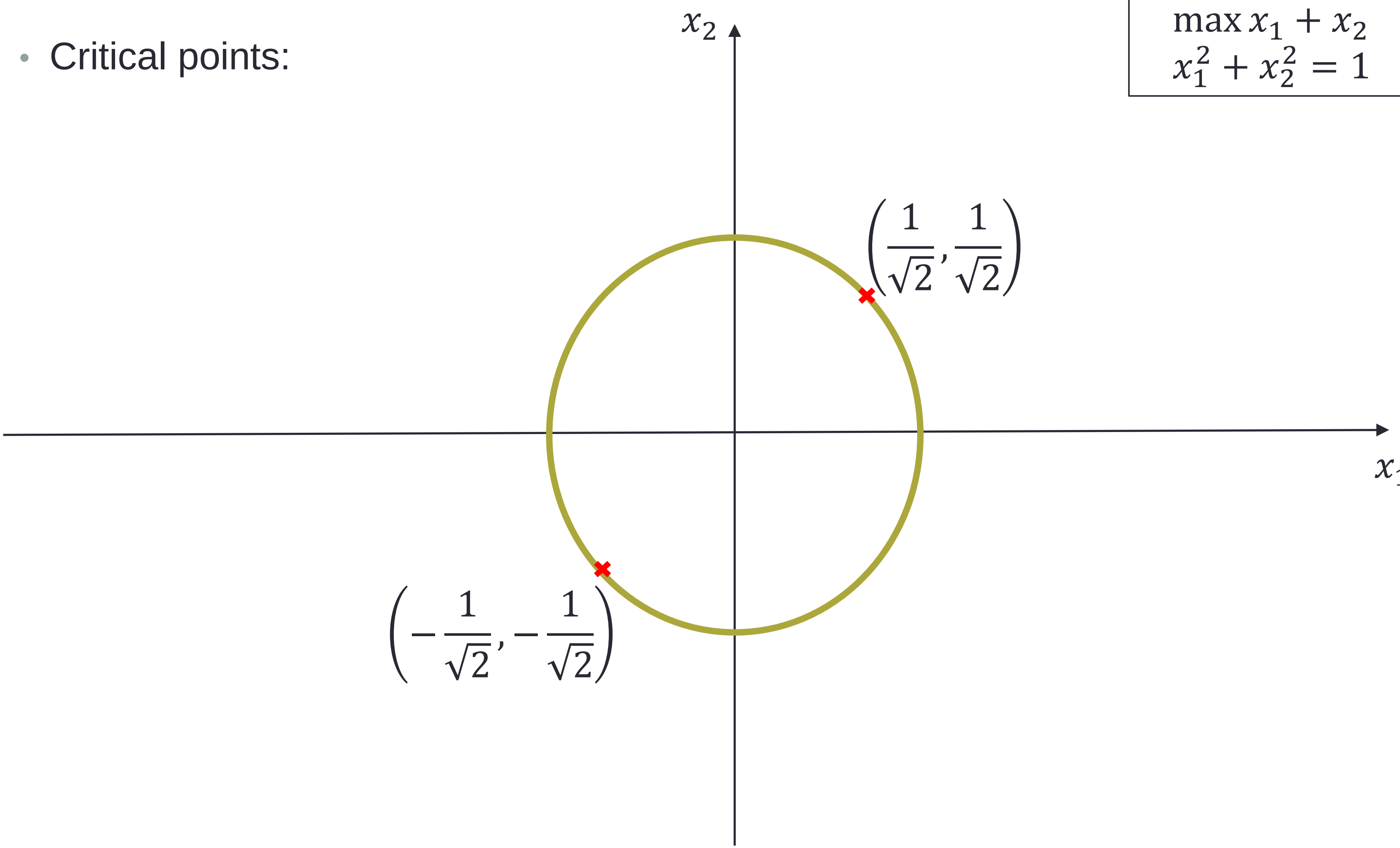
$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 = 1$$



Constrained Optimization with Equality Constraints

- Critical points:

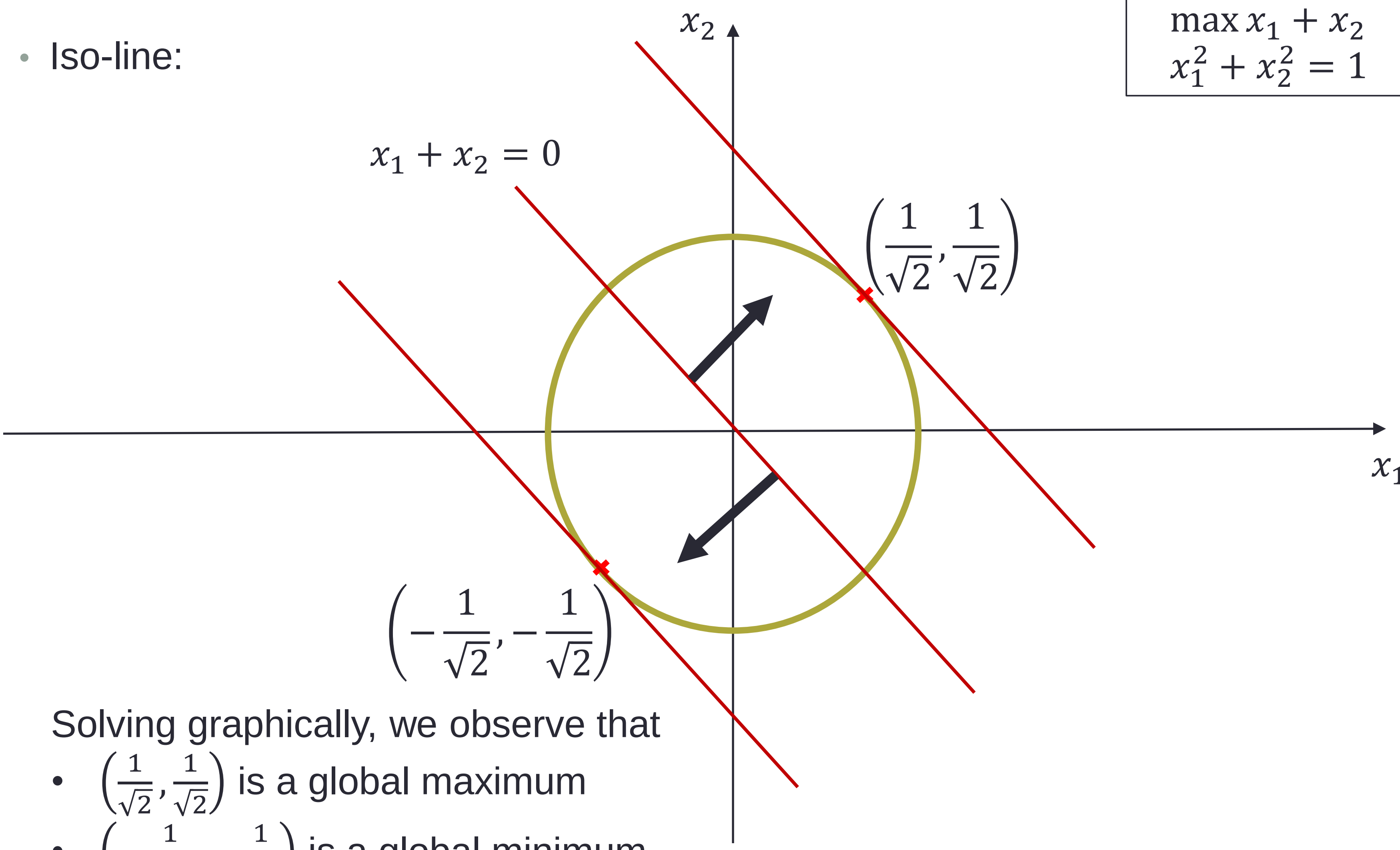
$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 = 1$$



Constrained Optimization with Equality Constraints

- Iso-line:

$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 = 1$$



Solving graphically, we observe that

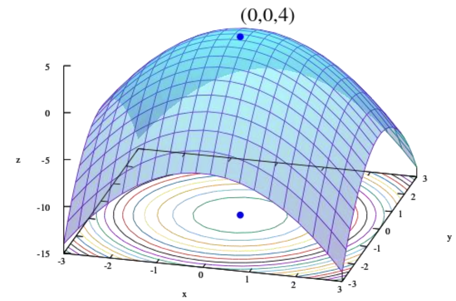
- $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ is a global maximum
- $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ is a global minimum

NLP: CONSTRAINED

With **inequality** constraints

- The KKT Conditions Method

Multi-var Constrained Opt



- Suppose we want to solve

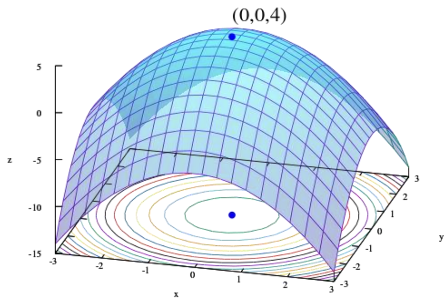
$$\max f(\boldsymbol{x}) \text{ subject to}$$

$$g_i(\boldsymbol{x}) \leq b_i \text{ for i} = 1, \ldots, m$$

where $f(\boldsymbol{x})$ and any of the $g_i(\boldsymbol{x})$ may be non-linear and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$

We will talk only about constrained optimization problems of the above form for the sake of following a convention

Multi-var Constrained Opt



- Suppose we want to solve

$$\max f(\boldsymbol{x}) \text{ subject to}$$

$$g_i(\boldsymbol{x}) \leq b_i \text{ for i} = 1, \ldots, m$$

where $f(\boldsymbol{x})$ and any of the $g_i(\boldsymbol{x})$ may be non-linear and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$

We will talk only about constrained optimization problems of the above form for the sake of following a convention

Karush

Kuhn

Tucker

KKT conditions

- Suppose we have an optimization problem of the following type:

$$\max f(\boldsymbol{x})$$
$$g_i(\boldsymbol{x}) \leq b_i \text{ for } i = 1, \ldots, m$$

where $f(\boldsymbol{x})$ and any of the $g_i(\boldsymbol{x})$ may be non-linear and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$

Theorem: If $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is a local or global optimum of the constrained problem, then there must be values $u = (u_1, u_2, \ldots, u_m)$ such that:

1. $\frac{\partial f(\boldsymbol{x})}{\partial x_j} - \sum_{i=1}^{m} u_i \frac{\partial g_i(\boldsymbol{x})}{\partial x_j} = 0$ for $j = 1,2, \ldots, n$
2. $g_i(\boldsymbol{x}) \leq b_i$ for $i = 1,2, \ldots, m$
3. $u_i \geq 0$ for $i = 1,2, \ldots, m$
4. $u_i(g_i(\boldsymbol{x}) - b_i) = 0$ for $i = 1,2, \ldots, m$

- Conditions 1, 2, 3, and 4 are called **Karush-Kuhn-Tucker (KKT) conditions**
- NOTE: Theorem also requires f and g to satisfy some “regularity conditions”, which will be satisfied by the functions encountered within the scope of this course

Constrained Optimization with Inequality Constraints (KKT conditions)

- Like the Lagrangian, we can use these conditions to identify critical points that could be local or global optima
- These points are called **KKT points**
- After computing the KKT points, pick the maximizer among them
  - Caution: Remember to verify if the picked point is a maximizer/minimizer
  - How to verify?
    1. Graphical method
    2. If the objection function f is concave and all constraint functions g_i are convex, then the KKT point that you found is a maximizer

Constrained Optimization with Inequality Constraints (KKT conditions)

- Consider the following constrained NLP:

$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 \leq 1$$
$$x_1 \leq \frac{1}{2}$$

$$f(\boldsymbol{x}) = x_1 + x_2$$
$$g_1(\boldsymbol{x}) = x_1^2 + x_2^2, \; b_1 = 1$$
$$g_2(\boldsymbol{x}) = x_1, \; b_2 = \frac{1}{2}$$

- The KKT conditions for this problem are:

1. $\frac{\partial f(\boldsymbol{x})}{\partial x_1} - u_1 \frac{\partial g_1(\boldsymbol{x})}{\partial x_1} - u_2 \frac{\partial g_2(\boldsymbol{x})}{\partial x_1} = 0$
2. $\frac{\partial f(\boldsymbol{x})}{\partial x_2} - u_1 \frac{\partial g_1(\boldsymbol{x})}{\partial x_2} - u_2 \frac{\partial g_2(\boldsymbol{x})}{\partial x_2} = 0$
3. $g_1(\boldsymbol{x}) \leq b_1$
4. $g_2(\boldsymbol{x}) \leq b_2$
5. $u_1(g_1(\boldsymbol{x}) - b_1) = 0$
6. $u_2(g_2(\boldsymbol{x}) - b_2) = 0$
7. $u_1, u_2 \geq 0$

1. $1 - 2u_1x_1 - u_2 = 0$
2. $1 - 2u_1x_2 = 0$
3. $x_1^2 + x_2^2 \leq 1$
4. $x_1 \leq \frac{1}{2}$
5. $u_1(x_1^2 + x_2^2 - 1) = 0$
6. $u_2\left(x_1 - \frac{1}{2}\right) = 0$
7. $u_1, u_2 \geq 0$

- Solving these equations provides the KKT points

Constrained Optimization with Inequality Constraints (KKT conditions)

1. $1 - 2u_1x_1 - u_2 = 0$
2. $1 - 2u_1x_2 = 0$
3. $x_1^2 + x_2^2 \leq 1$
4. $x_1 \leq \frac{1}{2}$
5. $u_1(x_1^2 + x_2^2 - 1) = 0$
6. $u_2\left(x_1 - \frac{1}{2}\right) = 0$
7. $u_1, u_2 \geq 0$

- To solve, try different cases for the u variables
- First choice: $u_1 = 0 \rightarrow$ Impossible (violates equation 2)
- What if $u_1 > 0$ and $u_2 = 0$
  - $x_1 = \frac{1}{2u_1}$ (by eqn 1)
  - $x_2 = \frac{1}{2u_1}$ (by eqn 2)
  - $u_1 = \frac{1}{\sqrt{2}}$ (by eqn 5)
  - Then $x_1 = \frac{1}{2u_1} = \frac{1}{\sqrt{2}}$ which contradicts constraint 4
- What if $u_1 > 0$ and $u_2 > 0$?
  - $x_1 = \frac{1}{2}$ (by eqn 6)
  - $x_2 = \pm\frac{\sqrt{3}}{2}$ (by eqn 5 with $u_1 > 0$)
  - If $x_2 = -\frac{\sqrt{3}}{2}$, $u_1 = -\frac{1}{\sqrt{3}}$ (by eqn 2), which contradicts $u_1 > 0$
  - If $x_2 = \frac{\sqrt{3}}{2}, u_1 = \frac{1}{\sqrt{3}}$ (by eqn 2)
  - $u_2 = 1 - \frac{1}{\sqrt{3}}$ (by eqn 1)
  - KKT point is $(x_1, x_2, u_1, u_2) = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}, \frac{1}{\sqrt{3}}, 1 - \frac{1}{\sqrt{3}}\right)$ →

Note: This is the unique KKT point. Need to verify if it is a maximizer or a minimizer

Constrained Optimization with Inequality Constraints (KKT conditions)

- Like the Lagrangian, we can use these conditions to identify critical points that could be local or global optima
- These points are called **KKT points**
- After computing the KKT points, pick the maximizer among them
  - Caution: Remember to verify if the picked point is a maximizer/minimizer
  - How to verify?
    1. Graphical method
    2. If the objection function f is concave and all constraint functions g_i are convex, then the KKT point that you found is a maximizer

Constrained Optimization with Inequality Constraints (KKT conditions)

- Consider the following constrained NLP:

$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 \leq 1$$
$$x_1 \leq \frac{1}{2}$$

$$f(\boldsymbol{x}) = x_1 + x_2$$
$$g_1(\boldsymbol{x}) = x_1^2 + x_2^2, \ b_1 = 1$$
$$g_2(\boldsymbol{x}) = x_1, \ b_2 = \frac{1}{2}$$

- Algebraic verification:
  - Check if f is concave and g_i is convex for every $i = 1, \ldots, m$

$\nabla^2 f = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ $\Rightarrow$ all eigenvalues are non-positive, so f is concave

$\nabla^2 g_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ $\Rightarrow$ all eigenvalues are non-negative, so g_1 is convex

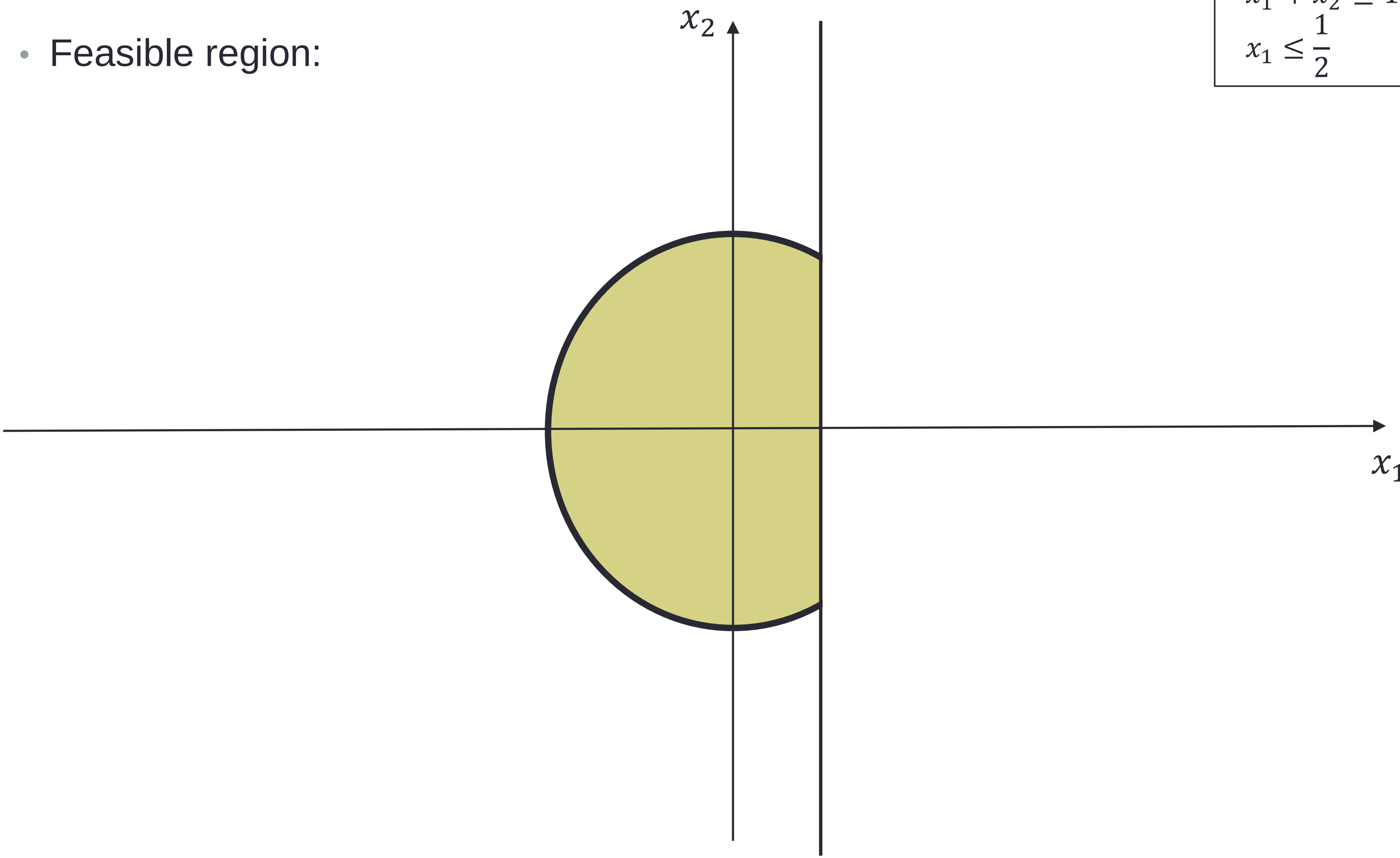
$\nabla^2 g_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ $\Rightarrow$ all eigenvalues are non-negative, so g_2 is convex

$\Rightarrow$ the unique KKT point is a maximizer

Constrained Optimization with Inequality Constraints (KKT conditions)

$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 \leq 1$$
$$x_1 \leq \frac{1}{2}$$

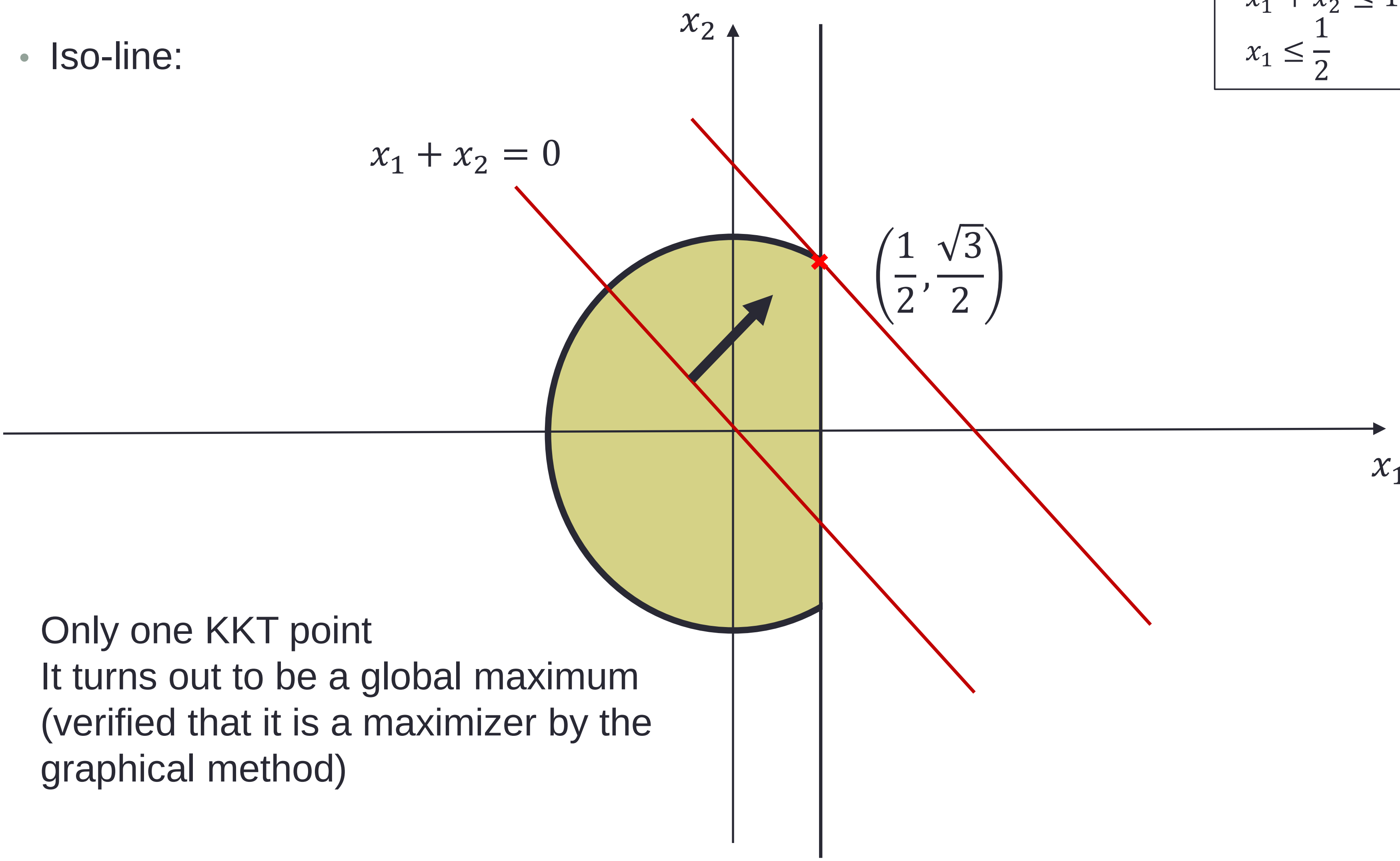
- Feasible region:



Constrained Optimization with Inequality Constraints (KKT conditions)

$$\max x_1 + x_2$$
$$x_1^2 + x_2^2 \leq 1$$
$$x_1 \leq \frac{1}{2}$$

- Iso-line:



Only one KKT point
It turns out to be a global maximum (verified that it is a maximizer by the graphical method)

KKT conditions: the significance of u_i

- Recall that one of the KKT conditions is

$$u_i(g_i(x) - b_i) = 0 \text{ for } i = 1,2, \dots, m$$

- This condition means that either:
  - The constraint is satisfied as equality, i.e., $g_i(x) = b_i$, and u_i may take a non-zero value
  - The constraint is a strict inequality, i.e., $g_i(x) < b_i$, and we must have $u_i = 0$
  - That is, this condition is complementary slackness condition!

- So, the value u_i is the **shadow price** of constraint i

Constrained Nonlinear Optimization

- We investigated two kinds of Constrained NLP:
  1. All constraints are equality constraints
     Solution method: Lagrange multipliers
  2. All constraints are inequality constraints
     Solution method: KKT conditions
  - Both these methods can be combined to consider both kinds of constraints together

- Finding a KKT point, or a stationary point for the Lagrangian function, *does not* guarantee that it is a local or global optimum without further analysis

> **Theorem.** If $f(\boldsymbol{x})$ is concave and each $g_i(\boldsymbol{x})$ is convex, then every KKT point (for a maximization problem) will be a global maximum

- In practice, it is very difficult to apply these methods to large problems, and consequently, nonlinear optimization is a vibrant area of research stemming from Machine Learning applications

NLP: ALGORITHMS

… where we see algorithms to solve single variable unconstrained nonlinear optimization problems

MOTIVATIONS

… why do we even need algorithms?
Why can't we use the theory?

Motivations for algorithm

- Suppose we want to solve $\max f(x)$
  where $f(x)$ is a differentiable concave function
- If we can find a point x^* such that $f'(x^*) = 0$, then it must be a global maximum

- We may not have an easy way to find such a stationary point
  - If $f(x)$ is a complicated function, there may not be an easy closed form expression for x^*
  - E.g., consider $f(x) = xe^{1-x} - x^6$
  - Is this function concave?

A complicated function

$$f(x) = xe^{1-x} - x^6$$

Is $f(x)$ concave?

$$f'(x) = e^{1-x}(1-x) - 6x^5$$
$$f''(x) = e^{1-x}(x-2) - 30x^4$$

Motivations for algorithm

- Suppose we want to solve $\max f(x)$

  where $f(x)$ is a differentiable concave function
- If we can find a point x^* such that $f'(x^*) = 0$, then it must be a global maximum

- We may not have an easy way to find such a stationary point
  - If $f(x)$ is a complicated function, there may not be an easy closed form expression for x^*
  - E.g., consider $f(x) = xe^{1-x} - x^6$
    - This function is concave

$$f'(x) = e^{1-x}(1-x) - 6x^5$$

Question: How do we find a root of $f'(x) = 0$?

We will find a value x that is numerically close to the value of the root