Assignment 1

- Due Feb. 4
- Online submission via CourSys
- Upload entire assignment in a single pdf (take photos if you wrote your solutions)
- Upload code separately via the code component

Convex Optimization II

CMPT 882

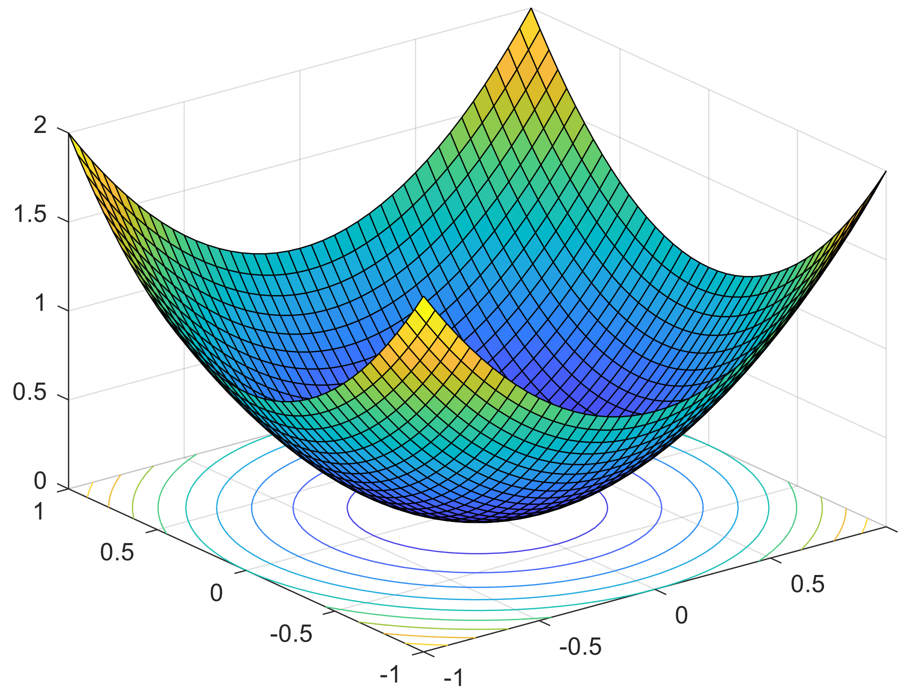
Feb. 1

Outline

- How to check if a function is convex

- Understand properties of optimal solutions

Convex Programs



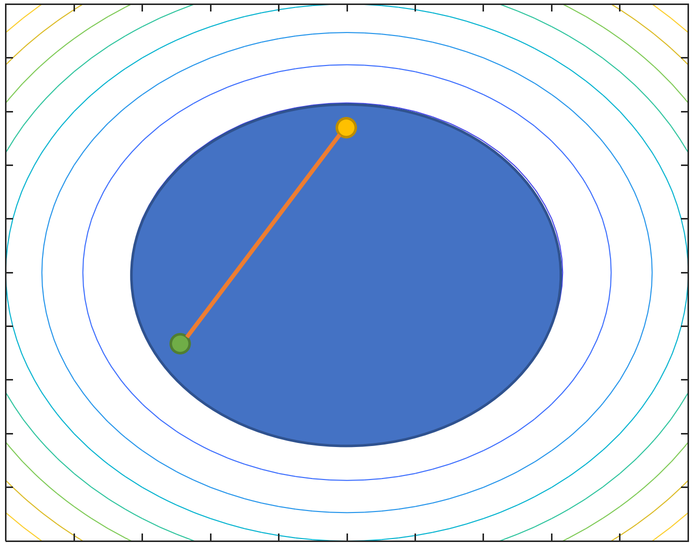
$$\text{minimize} \quad f(x)$$

$$\text{subject to} \quad g_i(x) \leq 0, i = 1, \ldots, n,$$

where $g_i(x)$ are convex

$$h_j^\top x = 0, j = 1, \ldots, m$$

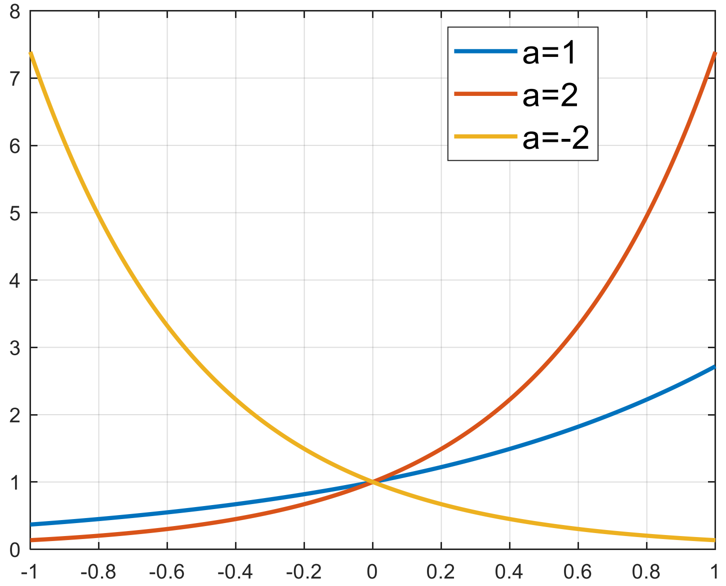
- Local optimum is global!
- Relatively easy to solve using simple algorithms
- When you see an optimization problem, first hope it's convex (although this is almost never true)
  - If an optimization problem is not convex, usually one can only hope for local optimum
- It is useful to recognize convex functions



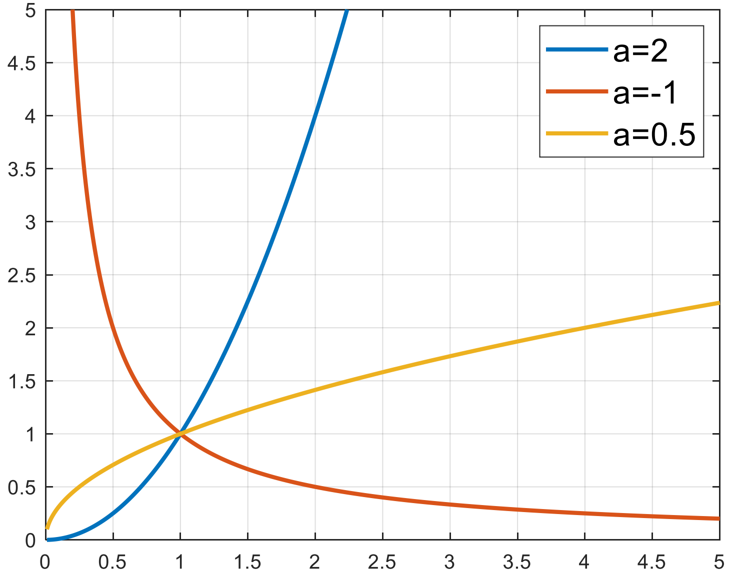
Common Convex Functions on $\mathbb{R}$

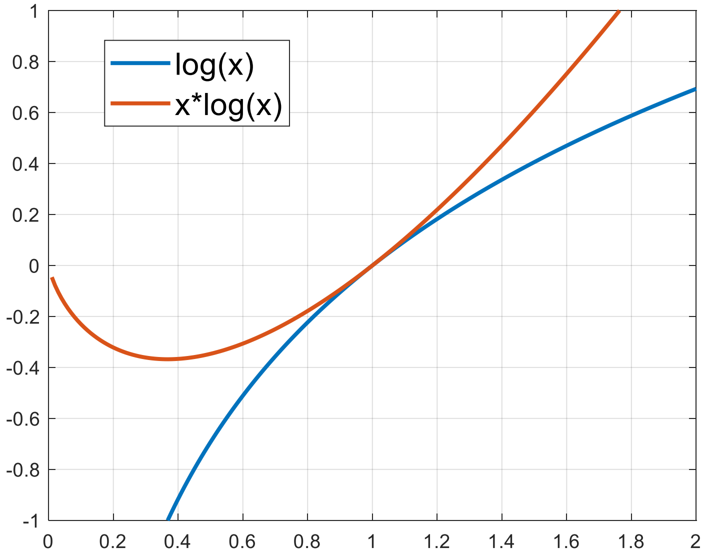
- $f(x) = e^{ax}$ is convex for all $x, a \in \mathbb{R}$
- $f(x) = x^a$ is convex on $x > 0$ if $a \geq 1$ or $a \leq 0$; concave if $0 < a < 1$
- $f(x) = \log x$ is concave
- $f(x) = x \log x$ is convex for $x > 0$ (or $x \geq 0$ if defined to be 0 when $x = 0$)

$f(x) = e^{ax}$



$f(x) = x^a$

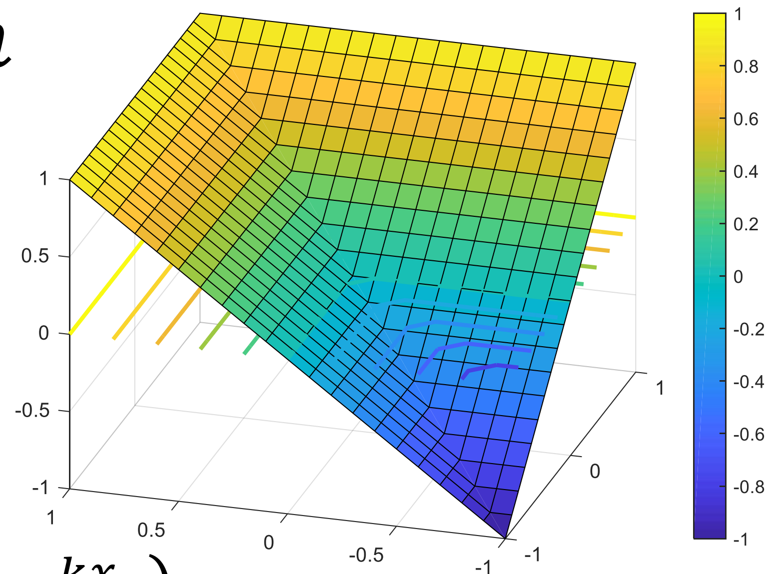




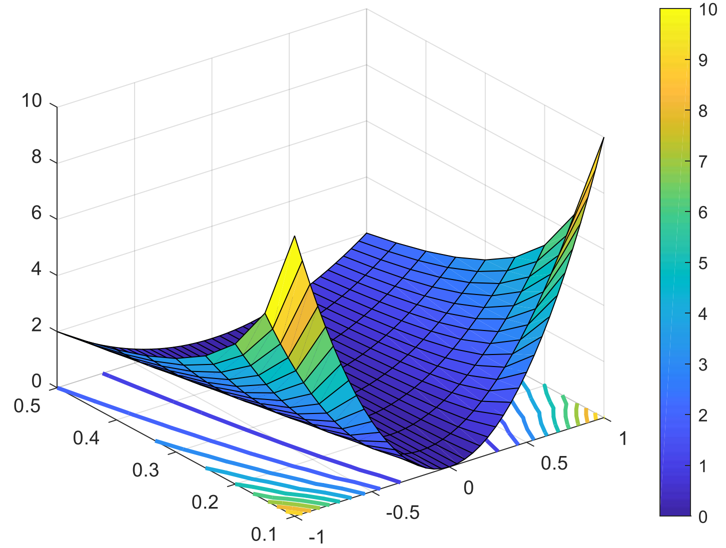
Common Convex Functions on $\mathbb{R}^n$

- $f(x) = Ax + b$ is convex for any A, b
- Every norm on $\mathbb{R}^n$ is convex
- $f(x) = \max(x_1, x_2, \dots, x_n)$ is convex
- $f(x) = \frac{x_1^2}{x_2}$ (for $x_2 > 0$)
- Log-sum-exp softmax: $f(x) = \frac{1}{k}\log\left(e^{kx_1} + e^{kx_2} + \cdots + e^{kx_n}\right)$
- Geometric mean: $f(x) = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}},\ x_i > 0$

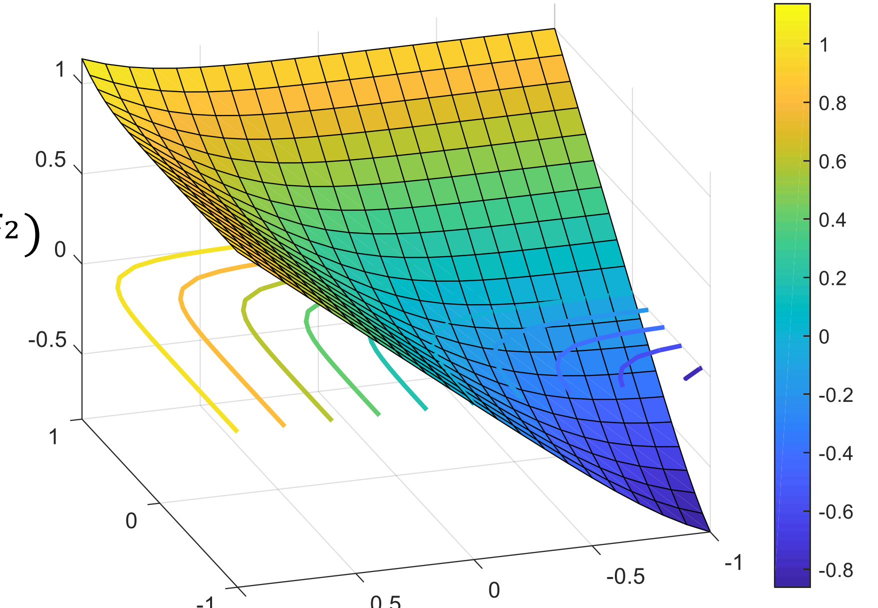
$f(x_1, x_2) = \max(x_1, x_2)$



$f(x) = \frac{x_1^2}{x_2}$



$f(x) = \frac{1}{5}\log(e^{5x_1} + e^{5x_2})$



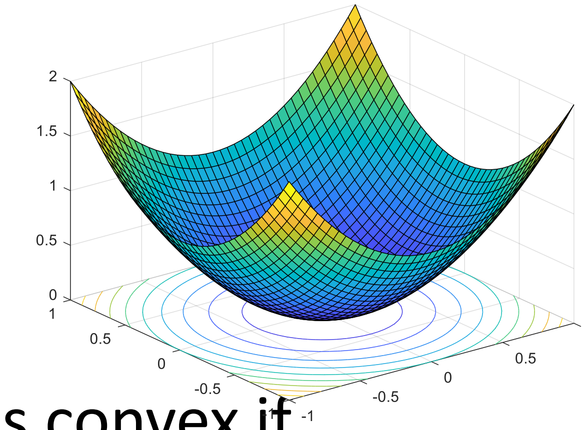
Operations that Preserve Convexity

- Non-negative weighted sum: $\sum_i w_i f_i(x)$ is convex if $f_i(x)$ are convex and $w_i \geq 0$
  - Example: $f(x) = ax^2 + bx^4 + cx^6$, where $a, b, c > 0$

- Composition with affine function: $g(x) = f(Ax + b)$ is convex if $f(x)$ is convex
  - Example: $f(\theta) = \|X\theta - Y\|_2^2$

- Point-wise maximum: $\max\big(f_1(x), f_2(x)\big)$

Operations that Preserve Convexity



- Minimization over a subset of variables: $g(y) := \min_z f(y,z)$ is convex if $f(y,z)$ is convex (jointly in (y,z))

- Perspective: $g(x,t) := tf\left(\frac{x}{t}\right)$, $t > 0$ is convex if $f(x)$ is convex
  - Example: $\frac{x_1^2}{x_2}$ is convex if $x_2 > 0$, because $f(x_1) = x_1^2$ is convex

- If $g_i: \mathbb{R}^n \to \mathbb{R}$ are convex, and $h: \mathbb{R}^k \to \mathbb{R}$ is convex and non-decreasing in each argument, then $h\big(g_1(x), g_2(x), \ldots, g_k(x)\big)$ is convex
  - Example: $\log(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$ is convex, because e^x is convex, and $\log x$ is convex and non-decreasing

How to check if a function is convex

- Use definition: $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$

Example 1:

- $f(x) = Ax + b, x \in \mathbb{R}^n$

$$\begin{aligned} f(\theta x + (1-\theta)y) &= A(\theta x + (1-\theta)y) + b \\ &= \theta Ax + (1-\theta)Ay + b \\ &= \theta Ax + (1-\theta)Ay + \theta b + (1-\theta)b \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

  - Equality!
  - This means f is also concave (i.e. $-f$ is convex)
  - Linear functions are both convex and concave

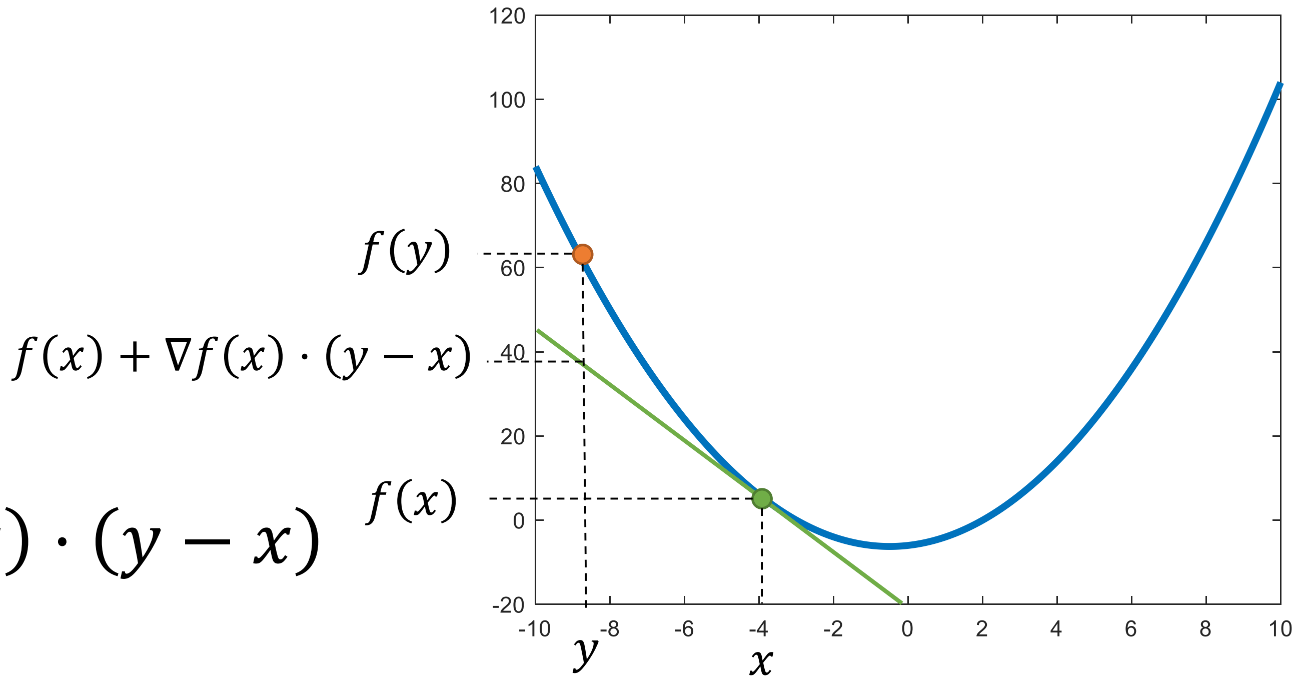
How to check if a function is convex

- Use definition: $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$

- Show $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ for differentiable functions

- Show $\nabla^2 f(x) \succcurlyeq 0$ for twice differentiable functions

Example 2:



- $f(x) = x^2 + x - 6$
- Method 1: show $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$
  - $\nabla f(x) = f'(x) = 2x + 1$

$$
\begin{aligned}
f(y) - f(x) + f'(x)(y - x) &= y^2 + y - 6 - [x^2 + x - 6 + (2x + 1)(y - x)] \\
&= y^2 + y - [x^2 + x + 2xy - 2x^2 + y - x] \\
&= y^2 + y - [-x^2 + 2xy + y] \\
&= y^2 + x^2 - 2xy \\
&= (x - y)^2 \geq 0
\end{aligned}
$$

- Method 2: show $\nabla^2 f(x) \geq 0$

$$\nabla^2 f(x) = f''(x) = 2 \geq 0$$

How to check if a function is convex

- Use definition: $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$
- Show $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ for differentiable functions
- Show $\nabla^2 f(x) \succcurlyeq 0$ for twice differentiable functions
- Show f is obtained from simple convex functions and operations that preserve convexity

Example 3:

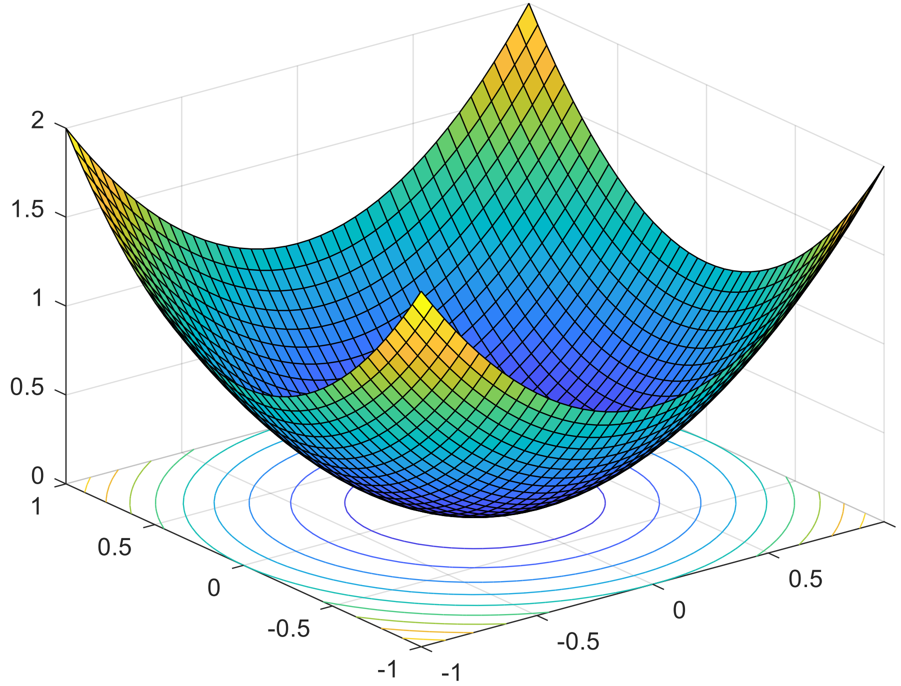
- $f(x) = \|Ax + b\|_2 + \lambda\|x\|_1$, A is a constant matrix, b is a constant vector, and $\lambda \geq 0$ is a constant scalar.
  - We know $\|x\|_1$ are $\|x\|_2$ are convex
    - All norms are convex
  - So, $\|Ax + b\|_2$ is convex, by the rule of affine composition
    - $g(x) = f(Ax + b)$ is convex if $f(x)$ is convex
  - Finally, $\|Ax + b\|_2 + \lambda\|x\|_1$ is convex, by the rule of non-negative weighted sum
    - $\sum_i w_i f_i(x)$ is convex if $f_i(x)$ are convex and $w_i \geq 0$

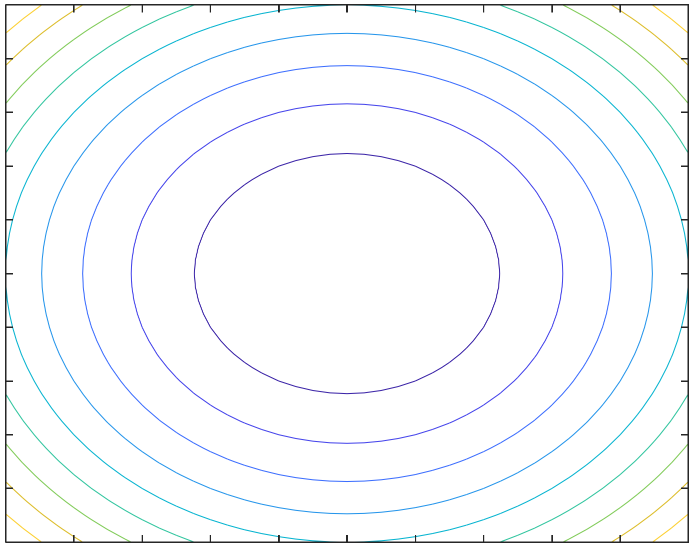
How to check if a function is convex

- Use definition: $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$
- Show $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ for differentiable functions
- Show $\nabla^2 f(x) \succcurlyeq 0$ for twice differentiable functions
- Show f is obtained from simple convex functions and operations that preserve convexity

Optimality Conditions for Convex Programs

- Unconstrained case: $\quad \text{minimize} \quad f(x)$
  - $\nabla f(x) = 0$





Optimality Conditions for Convex Programs

- Inequality constraints only:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, \dots, n \end{array}$$

- Penalty view point: penalize constraint violation
  - Lagrangian: $L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x)$, $\lambda_i \geq 0$

- Optimality conditions
  - Stationarity: $\nabla_x L(x^*, \lambda^*) = 0$
  - Primal feasibility: $g_i(x^*) \leq 0$
  - Dual feasibility: $\lambda^* \geq 0$
  - Complementary slackness: $\lambda_i^* g_i(x^*) = 0, \ i = 1, \dots, n$

Optimality Conditions for Convex Programs

- Stationarity: $\nabla_x L(x^*, \lambda^*) = 0$
  - Lagrangian:

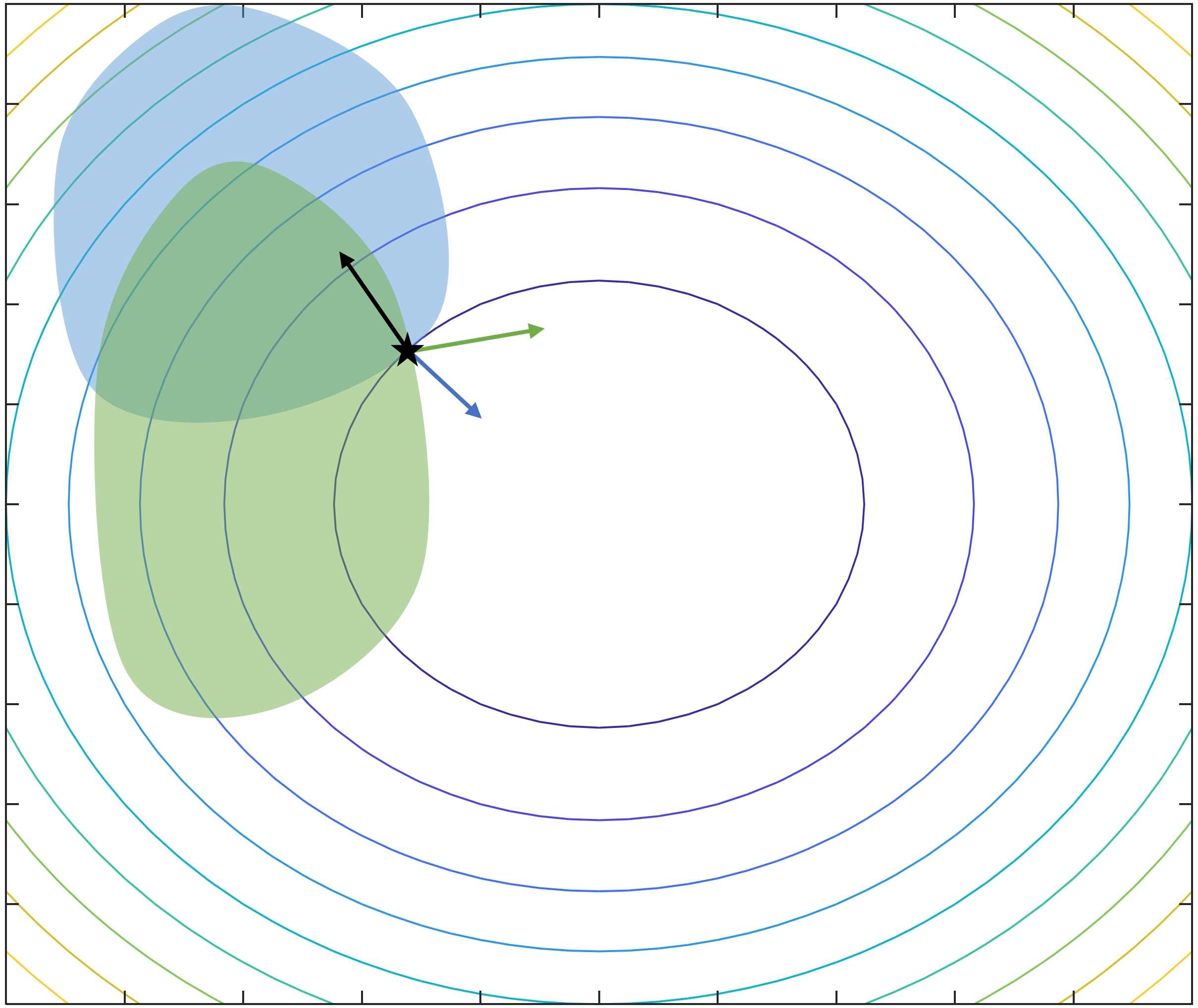
$$L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x), \qquad \lambda_i \geq 0$$

  - Take gradient and set to zero:

$$0 = \nabla f(x) + \sum_{i=1}^{n} \lambda_i \nabla g_i(x)$$

$$\nabla f(x) = -\sum_{i=1}^{n} \lambda_i \nabla g_i(x)$$

  - Since $\lambda_i \geq 0$, gradient of $f(x)$ must point “away” from gradients of active constraint functions



Optimality Conditions for Convex Programs

- Stationarity: $\nabla_x L(x^*, \lambda^*) = 0$
  - Lagrangian:

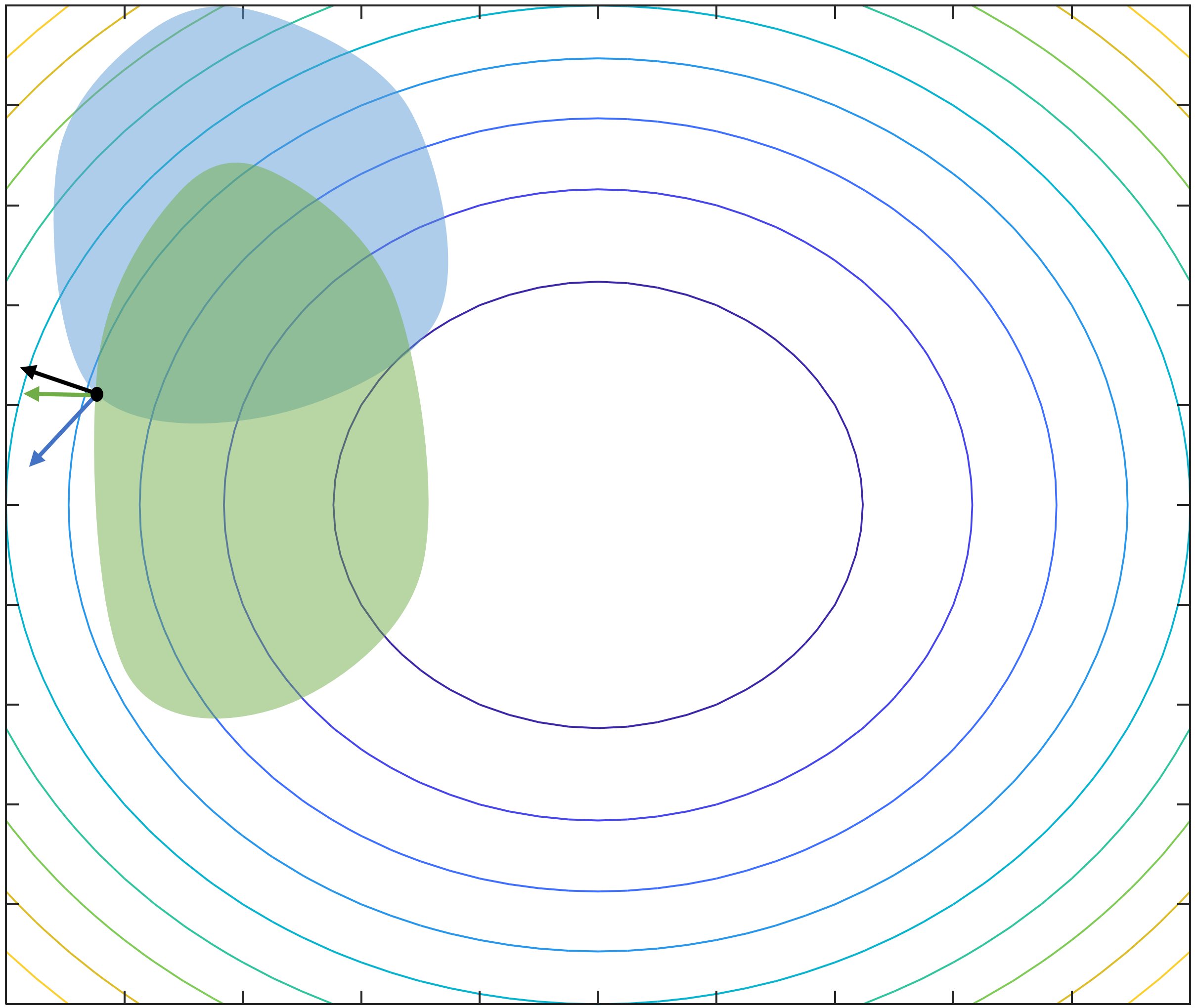
$$L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x), \qquad \lambda_i \geq 0$$

  - Take gradient and set to zero:

$$0 = \nabla f(x) + \sum_{i=1}^{n} \lambda_i \nabla g_i(x)$$

$$\nabla f(x) = -\sum_{i=1}^{n} \lambda_i \nabla g_i(x)$$

  - Since $\lambda_i \geq 0$, gradient of $f(x)$ must point "away" from gradients of active constraint functions



Optimality Conditions for Convex Programs

- Stationarity: $\nabla_x L(x^*, \lambda^*) = 0$
  - Lagrangian:

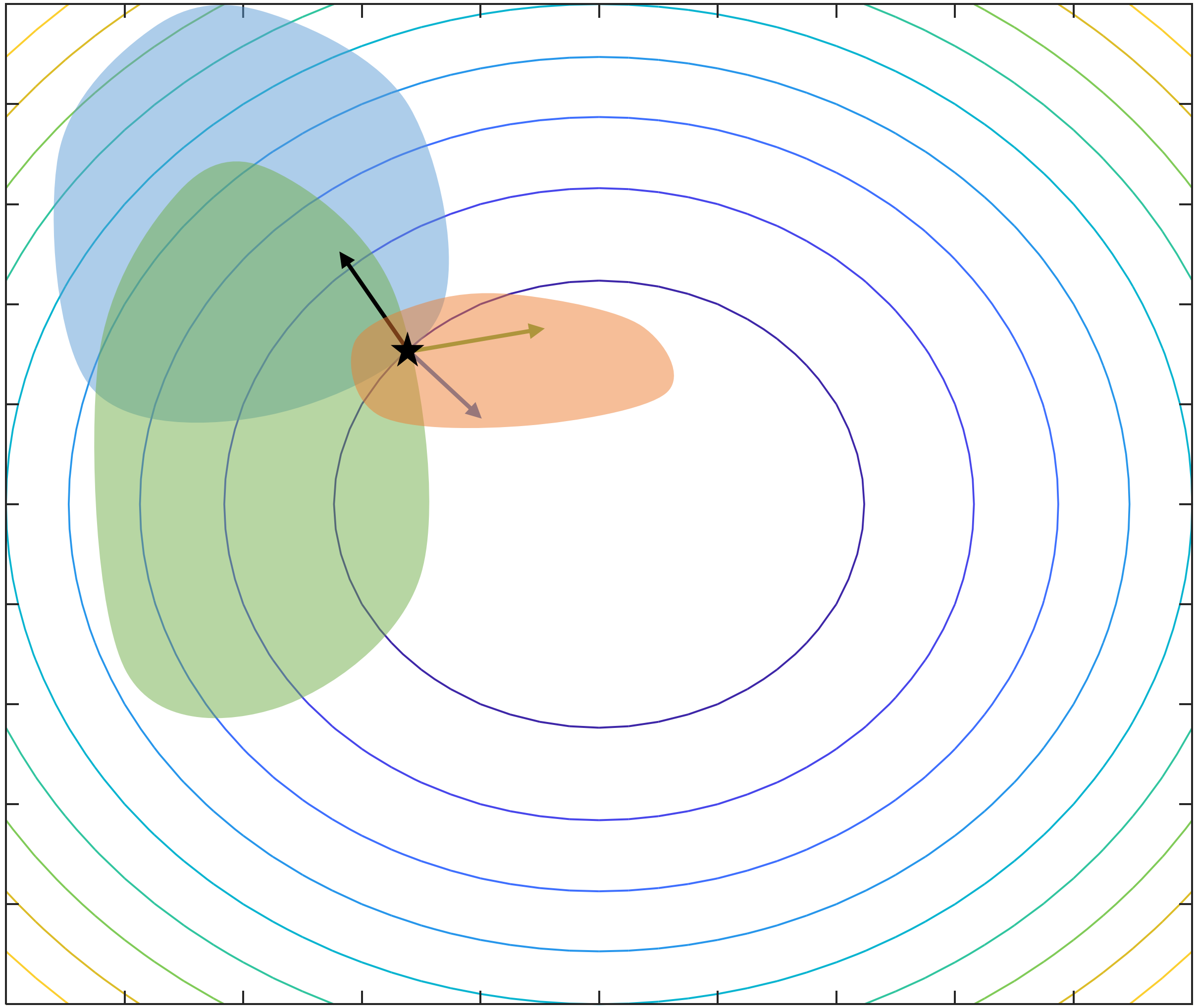
$$L(x,\lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x), \qquad \lambda_i \geq 0$$

  - Take gradient and set to zero:

$$0 = \nabla f(x) + \sum_{i=1}^{n} \lambda_i \nabla g_i(x)$$

$$\nabla f(x) = -\sum_{i=1}^{n} \lambda_i \nabla g_i(x)$$

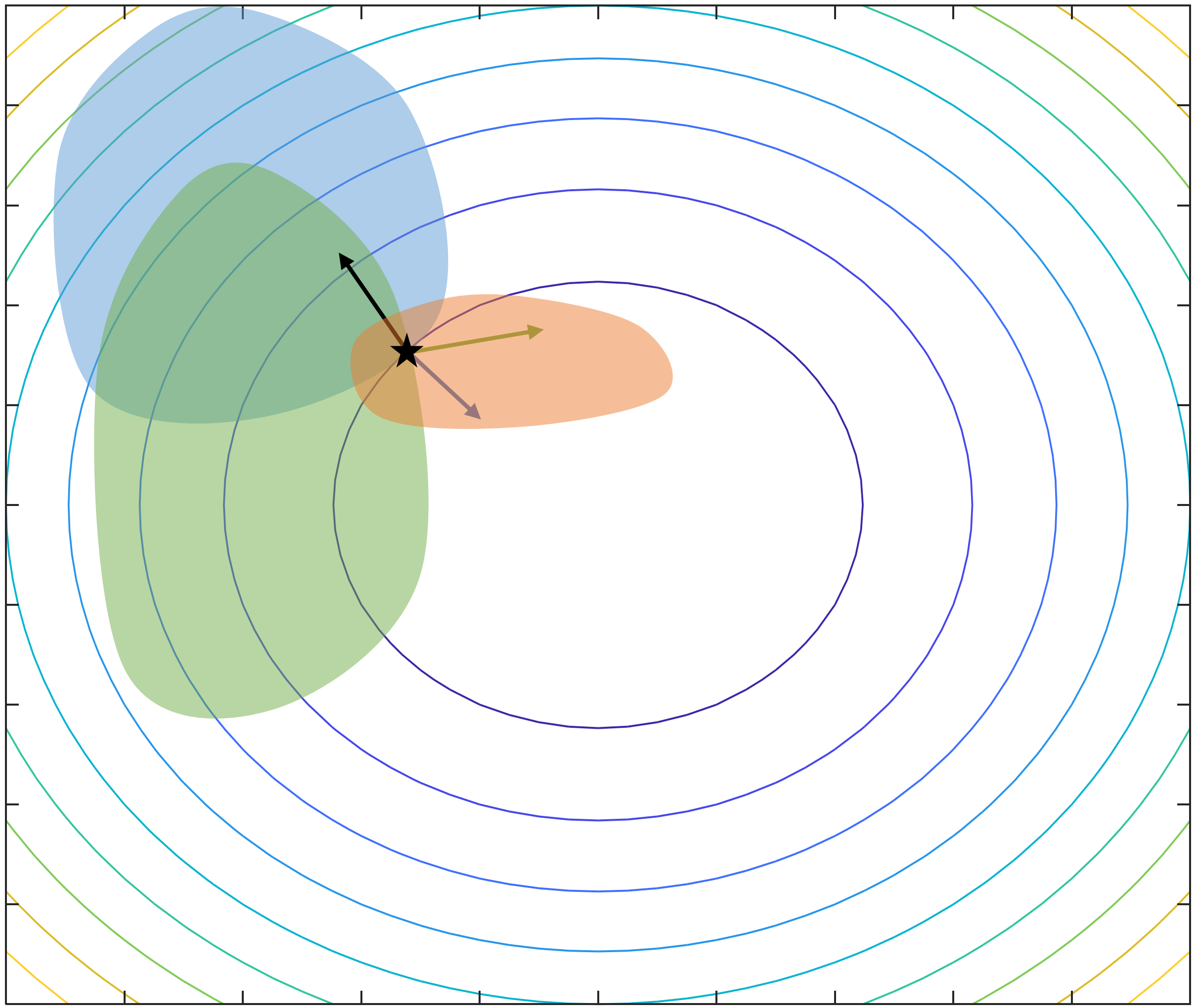
  - Since $\lambda_i \geq 0$, gradient of $f(x)$ must point "away" from gradients of **active** constraint functions



Optimality Conditions for Convex Programs

- Primal feasibility: $g_i(x^*) \leq 0$
  - Constraints must be satisfied

- Dual feasibility: $\lambda^* \geq 0$
  - Penalty view point

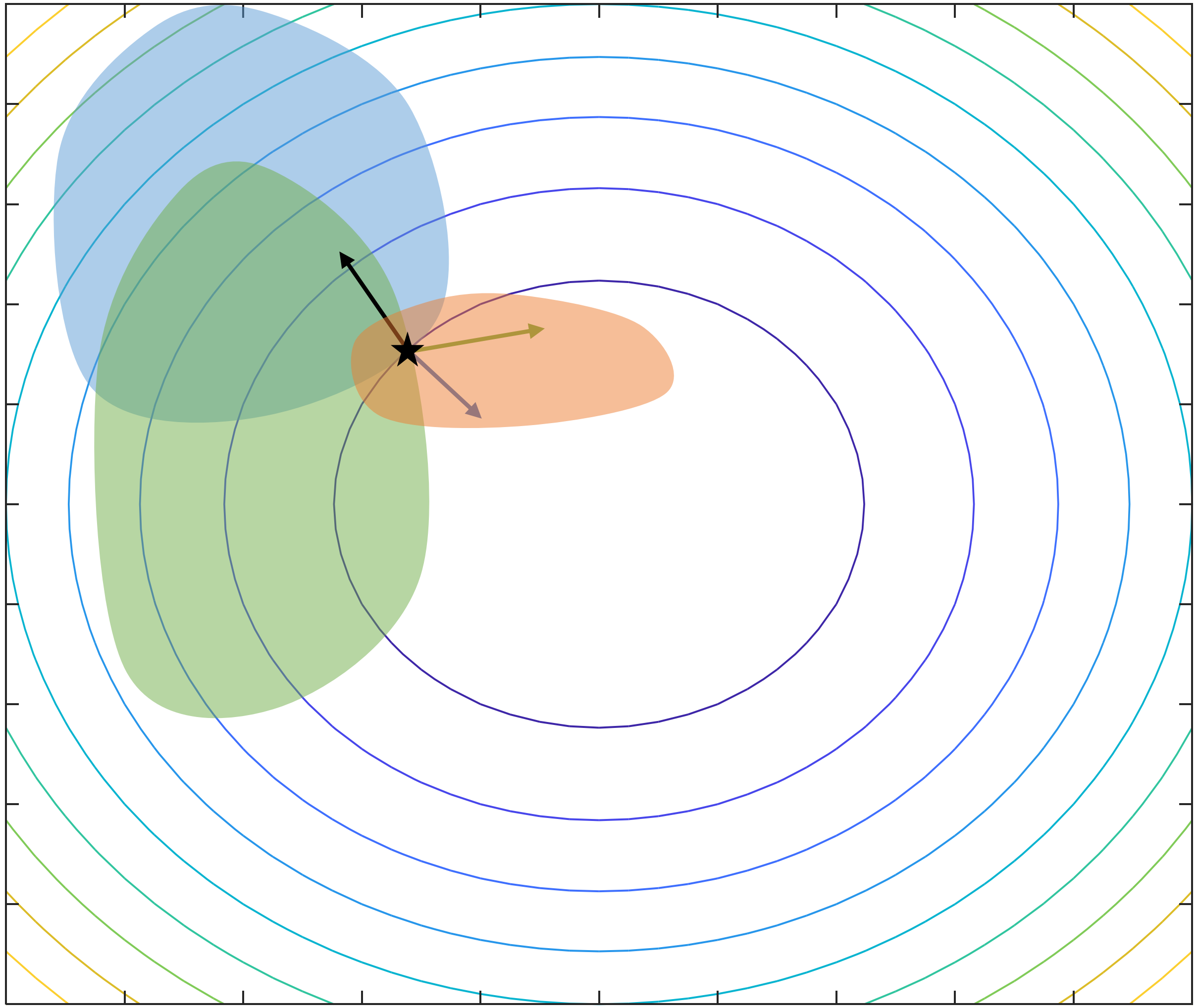


Optimality Conditions for Convex Programs

- Complementary slackness:
  $\lambda_i^* g_i(x^*) = 0, \; i = 1, \ldots, n$
  - Lagrangian:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x), \qquad \lambda_i \geq 0$$

  - If $g_i(x^*) < 0$, then the constraint is not active, so λ_i^* is set to 0 to not decrease the Lagrangian
  - If $g_i(x^*) = 0$, then the constraint is active, so λ_i^* is free to be positive



Optimality Conditions for Convex Programs

- Full optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \le 0, i = 1, \dots, n \\ & a_j^\top x = b_j, j = 1, \dots, m \end{array}$$

- Penalty view point:
  - Lagrangian: $L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + \sum_{j=1}^{m} \mu_j \left(a_j^\top x - b_j\right),\ \lambda_i \ge 0$
- Karush-Kuhn-Tucker (KKT) Conditions:
  - Stationarity $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$
  - Primal feasibility: $g_i(x^*) \le 0,\ a_i^\top x^* - b_i = 0$
  - Dual feasibility: $\lambda^* \ge 0$
  - Complementary slackness: $\lambda_i^* g_i(x^*) = 0,\ i = 1, \dots, n$

- Solve above systems of equations to obtain optimum