Convex and Affine Sets:

- $\mathcal{C}$ **affine if:** $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \mathcal{C} \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C},\ \lambda \in \mathbb{R}$
- $\mathcal{C}$ **convex if:** $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \mathcal{C} \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C},\ \lambda \in [0,1]$
- **Hyperplane:** $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^T\boldsymbol{x} = b\}$, $\boldsymbol{a} \neq \mathbf{0}$
- **Halfspace:** $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^T\boldsymbol{x} \leq b\}$, $\boldsymbol{a} \neq \mathbf{0}$
- **Polyhedron:** $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i^T\boldsymbol{x} \leq b_i, i = 1, \ldots, m,\ \boldsymbol{c}_j^T\boldsymbol{x} = d_j, j = 1, \ldots, p\}$
- **Separation Theorem:** $\mathcal{C}, \mathcal{D} \subseteq \mathbb{R}^n$ non-empty, convex with $\mathcal{C} \cap \mathcal{D} = \emptyset$.
$\Rightarrow \exists\ \boldsymbol{a} \in \mathbb{R}^n_{\neq 0}$ and $b \in \mathbb{R}$ such that $\boldsymbol{a}^T\boldsymbol{x} \leq b \leq \boldsymbol{a}^T\boldsymbol{y}\ \forall\ \boldsymbol{x} \in \mathcal{C}, \boldsymbol{y} \in \mathcal{D}$.
- **Supporting Hyperplane Theorem:** $\mathcal{C} \subseteq \mathbb{R}^n$ non-empty, convex.
$\Rightarrow \exists$ a supporting hyperplane at every boundary point of $\mathcal{C}$.

Convex Functions:

- f **[strictly] convex if:** $f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y})[<] \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$
- f **[strictly] concave if:** $-f$ [strictly] convex
- **Theorem (Restriction of a convex function to a line)** $f : \mathcal{C} \to \mathbb{R}$ is convex
$\Leftrightarrow g : \{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \mathcal{C}\} \to \mathbb{R},\ t \mapsto f(\boldsymbol{x} + t\boldsymbol{v})$ is convex (in t) for any $\boldsymbol{x} \in \mathcal{C}, \boldsymbol{v} \in \mathbb{R}^n$.
- **Theorem (First-order condition)** Differentiable f is convex
$\Leftrightarrow f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x})\ \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$.
- **Theorem (Second-order conditions)** f twice differentiable.
  1. f convex $\Leftrightarrow \nabla^2 f(\boldsymbol{x}) \geq 0\ \forall\ \boldsymbol{x} \in \mathcal{C}$.
  2. $\nabla^2 f(\boldsymbol{x}) > 0\ \forall\ \boldsymbol{x} \in \mathcal{C} \Rightarrow f$ strictly convex.
- **Theorem:** f convex $\Leftrightarrow$ epi(f) is convex.
- **Theorem (Minimizing a convex function over a convex set)** f convex and differentiable. Then, equivalent are
  1. $\boldsymbol{x}^*$ is a global minimum.
  2. $\boldsymbol{x}^*$ is a local minimum.
  3. $\boldsymbol{x}^*$ is a critical point, i.e., $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

Convex Optimization Problems:

- **Optimization problem in standard form:**

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && g_i(\boldsymbol{x}) \le 0, \quad i = 1, \dots, s \\ &&& h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, m \end{aligned}$$

- **Convex optimization problem in standard form:** f, g_i convex, $h_j(\boldsymbol{x}) = \boldsymbol{a}_j^T \boldsymbol{x} - b_j$

- **Linear program (LP):**

$$\begin{aligned} &\text{minimize} && \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d} \\ &\text{subject to} && \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h} \\ &&& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Equivalent convex problems:

- **Eliminating equality constraints:** $\boldsymbol{F}$ and $\boldsymbol{x}_0$ are such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \leftrightarrow \boldsymbol{x} = \boldsymbol{F}\boldsymbol{z} + \boldsymbol{x}_0$ for some $\boldsymbol{z}$.

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && g_i(\boldsymbol{x}) \le 0, i = 1, \dots, s \\ &&& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} &\text{minimize (over } \boldsymbol{z}) && f(\boldsymbol{F}\boldsymbol{z} + \boldsymbol{x}_0) \\ &\text{subject to} && g_i(\boldsymbol{F}\boldsymbol{z} + \boldsymbol{x}_0) \le 0, i = 1, \dots, s \end{aligned}$$

- **Introducing equality constraints:**

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{A}_0 \boldsymbol{x} + \boldsymbol{b}_0) \\ &\text{subject to} && g_i(\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i) \le 0, i = 1, \dots, s \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} &\text{minimize (over } \boldsymbol{x}, \boldsymbol{y}_i) && f(\boldsymbol{y}_0) \\ &\text{subject to} && g_i(\boldsymbol{y}_i) \le 0, i = 1, \dots, s \\ &&& \boldsymbol{y}_i = \boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i = 0, i = 1, \dots, s \end{aligned}$$

- **Introducing slack variables for linear inequalities:**

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && \boldsymbol{a}_i^T \boldsymbol{x} \le b_i, i = 1, \dots, s \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && \boldsymbol{a}_i^T \boldsymbol{x} + s_i = b_i, i = 1, \dots, s \\ &&& s_i \ge 0, i = 1, \dots, s \end{aligned}$$

- **Epigraph form:** Convex problem in standard form is equivalent to

$$\begin{aligned} &\text{minimize (over } \boldsymbol{x}, t) && t \\ &\text{subject to} && f(\boldsymbol{x}) - t \le 0 \\ &&& g_i(\boldsymbol{x}) \le 0, i = 1, \dots, s \\ &&& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

- **Minimizing over some variables:** $\tilde{f}(\boldsymbol{x}_1) = \inf_{\boldsymbol{x}_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2)$

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ &\text{subject to} && g_i(\boldsymbol{x}_1) \le 0, i = 1, \dots, s \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} &\text{minimize} && \tilde{f}(\boldsymbol{x}_1) \\ &\text{subject to} && g_i(\boldsymbol{x}_1) \le 0, i = 1, \dots, s \end{aligned}$$

Lagrangian Duality and KKT Conditions:

- **Lagrangian:** $L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{s} \lambda_i g_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j h_j(\boldsymbol{x})$
- **Lagrange dual function:** $L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$
- **Lower bound property:** $L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^*$ for any $\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu} \in \mathbb{R}^m$
- **Lagrange dual problem:**

$$\begin{array}{ll} \text{maximize} & L_D(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq 0. \end{array}$$

- **Theorem (Weak duality):** $d^* \leq p^*$
- **Strong duality:** $d^* = p^*$
- **Slater's constraint qualification:** Problem convex, and $\exists\ \boldsymbol{x} \in \text{int}\ \mathcal{D}$ with $g_i(\boldsymbol{x}) < 0\ \forall i, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
- **Theorem:** Slater's constraint qualification $\Rightarrow$ strong duality
- **KKT conditions:**
  1. Primal constraints: $g_i(\boldsymbol{x}) \leq 0, i = 1, \ldots, s,\ h_j(\boldsymbol{x}) = 0, j = 1, \ldots, m$
  2. Dual constraints: $\boldsymbol{\lambda} \geq \mathbf{0}$
  3. Complementary slackness: $\lambda_i g_i(\boldsymbol{x}) = 0, i = 1, \ldots, s$
  4. Gradient of Lagrangian with respect to $\boldsymbol{x}$ vanishes:

$$\nabla f(\boldsymbol{x}) + \sum_{i=1}^{s} \lambda_i \nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j \nabla h_j(\boldsymbol{x}) = \mathbf{0}$$

- **Theorem:** Consider a convex optimization problem with f, g_i, h_j differentiable, $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}}$ satisfying the KKT conditions $\Rightarrow \tilde{\boldsymbol{x}}, (\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\mu}})$ primal and dual optimal with zero duality gap.
- **Theorem:** Consider a convex optimization problem with f, g_i, h_j differentiable. Assume Slater's condition is satisfied. Then: $\boldsymbol{x}$ optimal $\Leftrightarrow \exists\ \boldsymbol{\lambda}, \boldsymbol{\mu}$ satisfying the KKT conditions.

Unconstrained Optimization:

- **Algorithm** General descent method
  given a starting point $\boldsymbol{x} \in \mathrm{dom} f$
  repeat

  1. Determine a descent direction $\Delta\boldsymbol{x}$.
  2. *Line search.* Choose step size t via exact or backtracking line search.
  3. *Update.* $\boldsymbol{x} := \boldsymbol{x} + t\Delta\boldsymbol{x}$.

  until stopping criterion is satisfied.

- **Exact line search:** $t = \mathrm{argmin}_{s>0} f(\boldsymbol{x} + s\Delta\boldsymbol{x})$.

- **Backtracking line search** (parameters $\alpha \in (0, \frac{1}{2}), \beta \in (0,1)$):
  starting at $t = 1$, repeat $t := \beta t$ until $f(\boldsymbol{x} + t\Delta\boldsymbol{x}) < f(\boldsymbol{x}) + \alpha t \nabla f(\boldsymbol{x})^T \Delta\boldsymbol{x}$.

- **Gradient descent method:** $\Delta\boldsymbol{x} = -\nabla f(\boldsymbol{x})$

- **Normalized steepest descent method:** $\Delta\boldsymbol{x} = \Delta\boldsymbol{x}_{nsd} = \mathrm{argmin}\{\nabla f(\boldsymbol{x})^T \boldsymbol{v} \mid \|\boldsymbol{v}\| = 1\}$

- **Algorithm** Newton's method
  given a starting point $\boldsymbol{x} \in \mathrm{dom} f$, tolerance $\varepsilon > 0$.
  repeat

  1. *Compute the Newton step and decrement.*

  $$\Delta\boldsymbol{x}_{nt} := -\nabla^2 f(\boldsymbol{x})^{-1} \nabla f(\boldsymbol{x}), \quad \lambda^2 := \nabla f(\boldsymbol{x})^T \nabla^2 f(\boldsymbol{x})^{-1} \nabla f(\boldsymbol{x})$$

  2. *Stopping criterion.* **quit** if $\lambda^2/2 \le \varepsilon$.
  3. *Line search.* Choose step size t via backtracking line search.
  4. *Update.* $\boldsymbol{x} := \boldsymbol{x} + t\Delta\boldsymbol{x}_{nt}$.

Constrained Optimization:

- **Equality Constrained Problems** The solution for the problem

  $$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array} \quad \text{may be achieved by solving} \begin{pmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x}_{nt} \\ \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} -\nabla f(\boldsymbol{x}) \\ \boldsymbol{0} \end{pmatrix}$$

  and executing the Newton method with step $\Delta\boldsymbol{x}_{nt}$.

- **Algorithm** Barrier method
  given a strictly feasible $\boldsymbol{x}, t := t^{(0)} > 0, \nu > 1$, tolerance $\varepsilon > 0$.
  repeat

  1. *Centering step.* Compute $\boldsymbol{x}^*(t)$ by minimizing $tf + \phi$, subject to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.
  2. *Update.* $\boldsymbol{x} := \boldsymbol{x}^*(t)$.
  3. *Stopping criterion.* **quit** if $s/t < \varepsilon$.
  4. *Increase t.* $t := \nu t$.