Chapter 2 Convexity

Notation

- $\|\mathbf{x}\|$: Euclidian norm, ℓ-2 norm.
- Cauchy-Schwarz ineq: $|\mathbf{u}^\top \mathbf{v}| \leqslant \|\mathbf{u}\| \cdot \|\mathbf{v}\|$
- spectual norm (2-norm) of matrix $A \in \mathbb{R}^{m\times n}$: $\|A\| := \max_{\mathbf{v}\in\mathbb{R}^d, \mathbf{v}\neq 0} \dfrac{\|A\mathbf{v}\|}{\|\mathbf{v}\|} = \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$
  - $\|A\mathbf{v}\| \leq \|A\|\|\mathbf{v}\|$

Convex Sets

- Defnition 2.7 A set $C \subseteq \mathbb{R}^d$ is convex if $\forall \mathbf{x}, \mathbf{y} \in C, \forall \lambda \in [0,1]$, $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C$.
- Observation 2.8 (Intersection) Let $C_i, i \in I$ be convex sets, then $C = \bigcap_{i\in I} C_i$ is a convex set.
- Theorem 2.9 (B-Lipschitz equivalence to Derivative) If (1) $f : \mathbf{dom}(f) \to \mathbb{R}^m$ is differentiable, (2) $X \subseteq \mathbf{dom}(f)$ is a non-empety and open convex set, then following equivalent:
  - (1) f is B-Lipschitz, $\|f(\mathbf{x}) - f(\mathbf{y})\| \leq B\|\mathbf{x}-\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in X$.
  - (2) f's Jacobian is bounded by B in spectual norm, $\|Df(\mathbf{x})\| \leq B, \forall \mathbf{x} \in X$.
  - Further, if X not open, then (2) implies (1).
  - Proof
    - (1)→(2)
      - By openness, $\forall \mathbf{x} \in X$, $\exists l$ s.t. ball $B(\mathbf{x}, l) \in X$,
      - By differentiability, $\forall \mathbf{x} \in X, \mathbf{v} \in B(\mathbf{x}, l)$, $f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + Df(\mathbf{x})\mathbf{v} + r(\mathbf{v})$, where $\lim_{\|\mathbf{v}\|\to 0} \dfrac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|} = 0$
      - By B-Lipschitz, $B\|\mathbf{v}\| \geq \|f(\mathbf{x}+\mathbf{v}) - f(\mathbf{x})\| = \|Df(\mathbf{x})\mathbf{v} + r(\mathbf{v})\| \geq \|Df(\mathbf{x})\mathbf{v}\| - \|r(\mathbf{v})\|$
      - Therefore $\dfrac{\|Df(\mathbf{x})\mathbf{v}\|}{\|\mathbf{v}\|} \leq B + \dfrac{\|r(\mathbf{v})\|}{\|\mathbf{v}\|}, \forall \mathbf{v} \in B(\mathbf{x}, l)$,
      - let $\mathbf{v}$ be the optimal direction where $\dfrac{\|Df(\mathbf{x})\mathbf{v}\|}{\|\mathbf{v}\|} = \|Df(\mathbf{x})\|$, and let its magnitude towards zero, then $\|Df(\mathbf{x})\| \leq B$.
    - (2)→(1), no need to assume open
      - For arbitary $\mathbf{x}, \mathbf{y} \in X \subseteq \mathbf{dom}(f), \mathbf{x} \neq \mathbf{y}$, define a scalar function for arbitary $\mathbf{z}$, $h(t) = \mathbf{z}^\top f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))$, $h'(t) = \mathbf{z}^\top Df(t) \times (\mathbf{y}-\mathbf{x})$
      - by mean value theorem, $\exists c \in (0,1)$ s.t. $h'(c) = h(1) - h(0)$, so $\mathbf{z}^\top Df(c) \times (\mathbf{y}-\mathbf{x}) = \mathbf{z}^\top (f(\mathbf{y}) - f(\mathbf{x}))$,
      - By Cauch-Schwarz ineq $\|\mathbf{z}^\top (f(\mathbf{y}) - f(\mathbf{x}))\| = \mathbf{z}^\top Df(c)(\mathbf{y}-\mathbf{x}) \leq \|\mathbf{z}\|\|Df(c)(\mathbf{y}-\mathbf{x})\|$
      - By spec-norm RHS $\leq |\mathbf{z}\|\|Df(c)\|\|(\mathbf{y}-\mathbf{x})\|$
      - By B-boundness of $\|Df\|$, RHS $\leq B\|\mathbf{z}\|\|(\mathbf{y}-\mathbf{x})\|$
      - Taking $\mathbf{z} = \dfrac{f(\mathbf{y}) - f(\mathbf{x})}{\|f(\mathbf{y}) - f(\mathbf{x})\|}$, LHS $= \|f(\mathbf{y}) - f(\mathbf{x})\| \leq$ RHS $= B\|(\mathbf{y}-\mathbf{x})\|$

Convex Functions

- Definition 2.10 A function f: $\mathbf{dom}(f) \to \mathbb{R}$ is *convex* if
  - (i) $\mathbf{dom}(f)$ is convex and
  - (ii) $\forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}(f), \lambda \in [0,1]$, $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$
- Definition of *Epigraph* $\mathrm{epi}(f) := \{(\mathbf{x}, \alpha) \in \mathbb{R}^{d+1} : \mathbf{x} \in \mathbf{dom}(f), \alpha \geq f(\mathbf{x})\}$
- Observation 2.11 f is a convex function if and only if $\mathrm{epi}(f)$ is a convex set.
  - Proof
    - $f \to \mathrm{epi}(f)$
      - $\forall (\mathbf{x}, \alpha), (\mathbf{y}, \beta) \in \mathrm{epi} f$, we know $\alpha \geq f(\mathrm{x}), \beta \geq f(\mathrm{y})$,
      - then $\forall \lambda \in [0,1]$, $\lambda\alpha + (1-\lambda)\beta \geq \lambda f(\mathrm{x}) + (1-\lambda) f(\mathrm{y}) \geq f(\lambda\mathrm{x} + (1-\lambda)\mathrm{y})$
      - then point $(\lambda\mathrm{x} + (1-\lambda)\mathrm{y}, \lambda\alpha + (1-\lambda)\beta) \in \mathrm{epi}(f)$
      - $\mathrm{epi}(f)$ convex
    - $\mathrm{epi}(f) \to f$
      - let $\alpha = f(\mathbf{x}), \beta = f(\mathbf{y})$, $\mathrm{epi}(f)$ convex means points $(\lambda\mathrm{x} + (1-\lambda)\mathrm{y}, \lambda f(\mathrm{x}) + (1-\lambda) f(\mathrm{y})) \in \mathrm{epi}(f)$,

- be def of $\mathrm{epi}(f)$, we have $\lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})\geq f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y})$

- Lemma 2.12 (Jensen's inequality) Let $f:\mathbb{R}^d\to\mathbb{R}$ be a convex function, $\mathbf{x}_1,\ldots,\mathbf{x}_m\in\mathbf{dom}(f)$, and $\lambda_1,\ldots,\lambda_m\in\mathbb{R}_+$ s.t. $\sum_{i=1}^m\lambda_i=1$, $f\left(\sum_{i=1}^m\lambda_i\mathbf{x}_i\right)\leq\sum_{i=1}^m\lambda_i f\left(\mathbf{x}_i\right)$
  - Proof Assume this holds for $m-1$, then $f\left(\sum_{i=1}^m\lambda_i\mathbf{x}_i\right)=f\left(\sum_{i=3}^m\lambda_i\mathbf{x}_i+(\lambda_1+\lambda_2)\frac{\lambda_1\mathbf{x}_1+\lambda_2\mathbf{x}_2}{\lambda_1+\lambda_2}\right)\leq(\lambda_1+\lambda_2)f(\frac{\lambda_1\mathbf{x}_1+\lambda_2\mathbf{x}_2}{\lambda_1+\lambda_2})+\sum_{i=3}^m\lambda_i f(\mathbf{x}_i)\leq\sum_{i=1}^m\lambda_i f(\mathbf{x}_i)$. $m=2$ holds as normal def of convexity
- Lemma 2.13 (Convexity and Continuity) If f is convex and suppose that $\mathbf{dom}(f)\subseteq\mathbb{R}^d$ is open, then f is continuous.
  - Proof (By hint from Homework)
    - Since $\mathbf{dom}(f)$ open, $\forall\mathbf{x}$, we can always find an open ball, and further a closed cube that $\mathbf{x}\in\otimes_{i=1}^d[l_i,r_i]$ and $\mathbf{x}$ being its center,
    - Any point $\mathbf{x}'$ in the cube can be written as normalized linear combination of corner. E.g. $\mathbf{x}'=(x_n,x_{-n})=\dfrac{r_n-x_n}{r_n-l_n}(l_n,x_{-n})+\dfrac{x_n-l_n}{r_n-l_n}(r_n,x_{-n})$
      - we can iteratively do this to every coordinate of $\mathbf{x}'$, until we get a normalized combination of 2^d corner
      - Therefore, $\forall\mathbf{x}'\in\otimes_{i=1}^d[l_i,r_i]$, $f(\mathbf{x}')\leq\sum_{i=1}^{2^d}\lambda_i' f(\mathbf{x}_i)\leq\max_i f(\mathbf{x}_i)$, $\mathbf{x}_i$ is corner.
    - If we do shrinkage over the cube by factor α_t, then each corner $\mathbf{x}_i$ becomes $(1-\alpha_t)\mathbf{x}+\alpha_t\mathbf{x}_i$
      - then $f((1-\alpha_t)\mathbf{x}+\alpha_t\mathbf{x}_i)\leq(1-\alpha_t)f(\mathbf{x})+\alpha_t f(\mathbf{x}_i)\to f(\mathbf{x})$ when $\alpha_t\to 0$
      - Therefore, we upper bound the value in the cube. $\forall\varepsilon,\exists\alpha_t$ s.t. $\max_i f(\mathbf{x}_{\alpha_t,i})-f(\mathbf{x})\leq\varepsilon$
    - Now need to lower bound it, if we already have $\forall\mathbf{x}'\in\otimes_{i=1}^d[l_{i,\alpha_t},r_{i,\alpha_t}]$, $f(\mathbf{x}')-f(\mathbf{x})\leq\varepsilon$.
      - If lower not bounded, e.g. $\exists\mathbf{y}\in\otimes_{i=1}^d[l_{i,\alpha_t},r_{i,\alpha_t}]$ s.t. $f(\mathbf{y})<f(\mathbf{x})-\varepsilon$,
      - by convexity $f(\mathbf{x})\leq(f(\mathbf{y})+f(2\mathbf{y}-\mathbf{x}))/2$, then $f(2\mathbf{y}-\mathbf{x})\geq 2f(\mathbf{x})-f(\mathbf{y})>f(\mathbf{x})+\varepsilon$.
      - contradictory, then in the cude, value also lower bounded by ε
    - We can also find a ball inside a cude, therefore f continuous.
- Lemma 2.14 (Counter Example in Infinite Dimension) $\exists$ vector space V and linear function f s.t. $\forall\mathbf{v}\in V$, f is discontinuous.
  - Example V be the polynomial function in $x\in[-1,1]$, distance measured by *supreme norm* $\|h\|_\infty:=\sup_{x\in[-1,1]}|h(x)\|$
    - consider the linear function mapping p to its derivative at $x=0$, $f:p(x)\to p'(0)$
    - consider zero polynomial $p_0(x)\equiv 0$, with $p_0'(0)=0$,
    - consider its neighbor $p_{n,k}(x)=\frac{1}{n}\sum_{i=0}^k(-1)^i\frac{(nx)^{2i+1}}{(2i+1)!}$ which is a finite expansion of $s_n(x)=\frac{1}{n}\sin(nx)=\frac{1}{n}\sum_{i=0}^\infty(-1)^i\frac{(nx)^{2i+1}}{(2i+1)!}$
    - since $\|p_{n,k}-s_n\|_\infty\to 0$ as $k\to\infty$ and $\|s_n\|_\infty\to 0$ as $n\to\infty$, this means $\|p_{n,k}\|\to 0$ as $n,k\to\infty$, on the other hand $f\left(p_{n,k}\right)=p_{n,k}'(0)=1$

Convexity Characterization

First Order

- Lemma 2.15 Suppose $\mathbf{dom}(f)$ open, f differentiable and $\forall\mathbf{x}$, $\nabla f(\mathbf{x})$ exists, then f is convex iff $\mathbf{dom}(f)$ convex, and $f(\mathbf{y})\geq f(\mathbf{x})+\nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})$
  - Proof
    - ($\to$), by convexity, for $t\in(0,1)$, $f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))\leq f(\mathbf{x})+t(f(\mathbf{y})-f(\mathbf{x}))$
      - $f(\mathbf{y})\geq f(\mathbf{x})+\frac{f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))-f(\mathbf{x})}{t}=f(\mathbf{x})+\nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})+\frac{r(t(\mathbf{y}-\mathbf{x}))}{t}$
      - Taking $t\to 0$ we have $f(\mathbf{y})\geq f(\mathbf{x})+\nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})$
    - ($\leftarrow$), for $\mathbf{z}:=\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\in\mathbf{dom}(f)$ we have $f(\mathbf{x})\geq f(\mathbf{z})+\nabla f(\mathbf{z})^\top(\mathbf{x}-\mathbf{z})$, and $f(\mathbf{y})\geq f(\mathbf{z})+\nabla f(\mathbf{z})^\top(\mathbf{y}-\mathbf{z})$
      - Weighted addition of two ineqs by factor $\{\lambda,1-\lambda\}$, we get convexity $\lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})\geq f(\mathbf{z})=f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y})$
- Lemma 2.16 (monotonicity of the gradient) Suppose $\mathbf{dom}(f)$ open, f differentiable. Then f is convex iff if dom(f) is convex and $\forall\mathbf{x},\mathbf{y}\in\mathbf{dom}(f),(\nabla f(\mathbf{y})-\nabla f(\mathbf{x}))^\top(\mathbf{y}-\mathbf{x})\geq 0$
  - Proof
    - ($\to$) If f convex, apply 2.15 to $\mathbf{x},\mathbf{y}$, and get $f(\mathbf{y})\geq f(\mathbf{x})+\nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})$ and $f(\mathbf{x})\geq f(\mathbf{y})+\nabla f(\mathbf{y})^\top(\mathbf{x}-\mathbf{y})$
      - add them togethor and get $0\geq(\nabla f(\mathbf{y})-\nabla f(\mathbf{x}))^\top(\mathbf{x}-\mathbf{y})$
    - ($\leftarrow$) Denote scalar function $h(t):=f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})),t\in[0,1]$,
      - so $h'(t)=\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))^\top(\mathbf{y}-\mathbf{x})$
      - setting $\mathbf{y}\leftarrow\mathbf{x}+t(\mathbf{y}-\mathbf{x})$ in monotonicity inequality, we have $h'(t)\geq h(0)\forall t\in[0,1]$
      - By mean value theorem, $\exists c$ s.t. $f(\mathbf{y})=h(1)=h(0)+h'(c)\geq h(0)+h'(0)=f(\mathbf{x})+\nabla f(\mathbf{y})^\top(\mathbf{y}-\mathbf{x})$, convexity

Second Order

- Lemma 2.17 Suppose $\mathbf{dom}(f)$ open and f twice differentiable, and hessian $\nabla f = (\partial_{ij} f)$ exists and symmetric. Then f is convex iff $\mathbf{dom}(f)$ convex and $\forall \mathbf{x} \in \mathbf{dom}(f)$, $\nabla^2 f(\mathbf{x}) \succeq 0$.
  - Proof
    - ($\to$) Denote $\mathbf{v} = \mathbf{x} - \mathbf{y}$, again define $h(t \in [[0,1]) := f(\mathbf{x} + t\mathbf{v})$.
      - We have $h'(t) = \nabla f(\mathbf{x} + t\mathbf{v})^\top \mathbf{v}$ and $h''(t) = \mathbf{v}^\top \nabla^2 f(\mathbf{x} + t\mathbf{v})\mathbf{v}$.
      - Since $\mathbf{dom}(f)$ open, $\forall \mathbf{x}, \exists \mathcal{U}(x, \delta) \in \mathbf{dom}(f)$, then we can set $\mathbf{v}$ to be arbitary on $\|\mathbf{v}\| = 1$.
      - Since f convex, h is also convex, by *Lemma 2.16* we have $(h'(\delta) - h'(0))\delta \geq 0 \Rightarrow h'(\delta) - h'(0))/\delta \geq 0$
      - Taking limit of $\delta \to 0$ we have $h'(\delta) - h'(0))/\delta \to h''(0) \geq 0$
      - This holds for arbitary $\|\mathbf{v}\| = 1$, therefore f positive semi-definite.
    - ($\leftarrow$) Assume $\nabla^2 f(\mathbf{x}) \succeq 0$, then we have $h'(t) \geq 0$ for arbitary $t \in [0,1]$ and $\mathrm{x}, \mathrm{y} \in \mathbf{dom}(f)$
      - Then $(\nabla f(\mathrm{y}) - \nabla f(\mathrm{x}))^\top (\mathrm{y} - \mathrm{x}) = h'(1) - h'(0) = \int_0^1 h''(t)dt \geq 0$
  - PS: Hessians of a twice continuously differentiable function are symmetric is a classical result known as the *Schwarz theorem*. If f twice differentiable, symmetry already holds. If f is only twice *partially* differentiable, we may have non-symmetric Hessians.

Operations Preserving Convexity

- Lemma 2.18
  - (i) Let $f_1, f_2, \ldots, f_m$ be convex functions and $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}_+$. Then (1) $\mathbf{max}_{i=1}^m f_i$ (2) $f := \sum_{i=1}^m \lambda_i f_i$ convex on $\mathbf{dom}(f) := \bigcap_{i=1}^m \mathbf{dom}(f_i)$.
  - (ii) Let f convex on $\mathbf{dom}(f) \subseteq \mathbb{R}^d$, $g : \mathbb{R}^m \to \mathbb{R}^d, \mathbf{x} \to A\mathbf{x} + \mathbf{b}$ be an affine function. Then $f \circ g$ convex on $\mathbf{dom}(f \circ g) := \{\mathbf{x} \in \mathbb{R}^m : g(\mathbf{x}) \in \mathbf{dom}(f)\}$.
- Given f, g convex, $f \circ g$ may be non-convex, example: $f(x) = x^2$, $g(x) = x^2 - 1$, $(f \circ g)(x) = x^4 - 2x^2 + 1$, $(f \circ g)(-1) = (f \circ g)(1) = 0$ and $(f \circ g)(0) = 1$.

Minimizer Condition

- Definition 2.19 A *local minimum* of $f : \mathrm{dom}(f) \to \mathbb{R}$ is a point x s.t. $\exists \varepsilon > 0, \forall \mathbf{y} \in \mathrm{dom}(f)$ s.t. $\|\mathbf{y} - \mathbf{x}\| < \varepsilon$, we have $f(\mathbf{x}) \leq f(\mathbf{y})$
- Lemma 2.20 (Global minimum) Let $\mathbf{x}^\star$ be a local minimum of a convex function $f : \mathrm{dom}(f) \to \mathbb{R}$. Then $\mathbf{x}^\star$ is a global minimum, if $\forall \mathbf{y} \in \mathrm{dom}(f), f(\mathbf{x}^\star) \leq f(\mathbf{y})$.
  - not all convex function have global minimum.
- Lemma 2.21 (Zero grad -> Global minimum) Suppose that $f : \mathrm{dom}(f) \to \mathbb{R}$ is convex and differentiable over an open domain $\mathrm{dom}(f) \subseteq \mathbb{R}^d$. Let $\mathbf{x} \in \mathbf{dom}(f)$. If $\nabla f(\mathbf{x}) = \mathbf{0}$, then $\mathbf{x}$ is a global minimum.
  - Proof $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) = f(\mathbf{x})$.
- Lemma 2.22 (Global minimum -> zero grad) Suppose that $f : \mathrm{dom}(f) \to \mathbb{R}$ is convex and differentiable over an open domain $\mathrm{dom}(f) \subseteq \mathbb{R}^d$. If $\mathbf{x}$ is a global minimum, then $\nabla f(\mathbf{x}) = \mathbf{0}$.

Strictly convex function

Note: not strong convex function

- Definition 2.23 (Strict convexity) A function $f : \mathrm{dom}(f) \to \mathbb{R}$ is *strictly convex* if (i) $\mathrm{dom}(f)$ is convex and (ii) $\forall \mathbf{x} \neq \mathbf{y} \in \mathbf{dom}(f)$ and $\forall \lambda \in (0,1)$, $f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y})$.
- Lemma 2.24 (Positive definite -> strict convexity) Suppose that $\mathrm{dom}(f)$ is open and that f is twice continuously differentiable. If $\forall \mathbf{x} \in \mathbf{dom}(f)$ the Hessian $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$, then f is strictly convex.
  - Proof similar to 2.17, only that $\mathbf{v} \neq 0$ so $\geq \to >$.
  - Converse (Strict convexity -> posi definite) is false, counter example: $f(x) = x^4$.
- Lemma 2.25 (Strict convexity -> Unique minimum). Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be strictly convex. Then f has at most one global minimum.
  - If $\mathbf{x}^\star \neq \mathbf{y}^\star$ both global minimum, then by strict convexity $\mathbf{z} = \frac{1}{2}\mathbf{x}^\star + \frac{1}{2}\mathbf{y}^\star$ have $f(\mathbf{z}) < \frac{1}{2} f_{\min} + \frac{1}{2} f_{\min} = f_{\min}$.

Constrained Minimization

- Definition 2.26 (minimizer on subset) Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be convex and let $X \subseteq \mathbf{dom}(f)$ be a convex set. A point $\mathbf{x} \in X$ is a *minimizer* of f over X if $\forall \mathbf{y} \in X, f(\mathbf{x}) \leq f(\mathbf{y})$.
  - Subset can be lower dimension embedded to $\mathrm{dom}(f)$.
- Lemma 2.27 (First order condition, variational inequality) Suppose that $f : \mathrm{dom}(f) \to \mathbb{R}$ is convex and differentiable over an open domain $\mathrm{dom}(f) \subset \mathbb{R}^d$, and let $X \subseteq \mathrm{dom}(f)$ be a convex set. $\mathbf{x}^\star \in X$ is a minimizer of f over X iff $\forall \mathbf{x} \in X, \nabla f(\mathbf{x}^\star)^\top (\mathbf{x} - \mathbf{x}^\star) \geq 0$.

- Proof
  - ($\rightarrow$) Suppose $\mathbf{x}^\star$ is minimizer, then for all $\mathbf{x} \in X$, $f(\mathbf{x}) - f(\mathbf{x}^\star) \geq 0$.
    - If $\exists \mathbf{x}'$ s.t. $\nabla f(\mathbf{x}^\star)^\top (\mathbf{x}' - \mathbf{x}^\star) < 0$, then $h(t \in [0,1]) := f(\mathbf{x}^\star) + \nabla f(\mathbf{x}^\star)^\top (\mathbf{x}' - \mathbf{x}^\star) < f(\mathbf{x}^\star)$
    - by Tayler theorem, we have $f(\mathbf{x}^\star + t(\mathbf{x}' - \mathbf{x}^\star)) < f(\mathbf{x}^\star)$, contradict to convexity in $\mathbf{dom}(f)$.
  - ($\leftarrow$) Suppose $\forall \mathbf{x} \in X, \nabla f(\mathbf{x}^\star)^\top (\mathbf{x} - \mathbf{x}^\star) \geq 0$,
    - by convexity in $\mathbf{dom}(f)$ we have, $\forall \mathbf{x} \in X$, $f(\mathbf{x}) \geq f(\mathbf{x}^\star + t(\mathbf{x} - \mathbf{x}^\star)) \geq f(\mathbf{x}^\star) + t\nabla f(\mathbf{x}^\star)^\top (\mathbf{x} - \mathbf{x}^\star) \geq f(\mathbf{x}^\star)$

Existence of minimizer

Sublevel sets, Weierstrass Theorem

- Definition 2.28 Let $f : \mathbb{R}^d \to \mathbb{R}, \alpha \in \mathbb{R}$. Set $f^{\leq \alpha} := \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \leq \alpha\}$ is the α-*sublevel set* of f.
  - If f convex, $f^{\leq \alpha}$ is convex set.
  - If f continuous (implied by convexity and finit dim), $f^{\leq \alpha}$ is closed.
- Theorem 2.29 (Weierstrass) Let $f : \mathbb{R}^d \to \mathbb{R}, \alpha \in \mathbb{R}$ be a continuous function, and suppose there is a nonempty and bounded sublevel set $f^{\leq \alpha}$. Then f has a global minimum.
  - Proof
    - since $(-\infty, \alpha]$ is closed, by continuity of f, its pre-image $f^{\leq \alpha}$ is closed.
    - From the fact that continuous function attains minimum over closed and bounded (=compact) set. We know f have minimum $\mathbf{x}^\star \in f^{\leq \alpha}$.
    - This is global minimum. Otherwise, if $\exists \mathbf{x}' \notin f^{\leq \alpha}$, then $f(\mathbf{x}') > \alpha \geq f(\mathbf{x}^\star)$.

Recession cone and lineality space

Examples like $f(x) = e^x$ is convex but does not have a global minimum.

- Definition 2.30 (unbounded in direction) Let $C \subseteq \mathbb{R}^d$ be a convex set. Then $\mathbf{y} \in \mathbf{R}^d$ is a *direction of recession* of C if $\exists \mathbf{x} \in C$ and $\forall \lambda \geq 0 \; \mathbf{x} + \lambda \mathbf{y} \in C$.
- Lemma 2.31 (Equivalance for existance and arbitry) Let $C \subseteq \mathbf{R}^d$ be a nonempty closed convex set, and let $\mathbf{y} \in \mathbb{R}^d$. $\forall \lambda \geq 0, \exists \mathbf{x} \in C : \mathbf{x} + \lambda \mathbf{y} \in C \Leftrightarrow \forall \lambda \geq 0, \forall \mathbf{x} \in C : \mathbf{x} + \lambda \mathbf{y} \in C$
  - Proof
    - ($\leftarrow$) is obvious. ($\rightarrow$): We already have $\mathbf{x}_0$ s.t. $\forall \lambda \geq 0, \mathbf{x}_0 + \lambda \mathbf{y} \in C$.
      - For arbitary $\mathbf{x} \in C$, let $\mathbf{z} = \lambda \mathbf{y}$, define $\mathbf{w}_k := \mathbf{x}_0 + k\mathbf{z} \in C$ by our assumption,
      - Define $\mathbf{z}_k := \frac{1}{k}(\mathbf{w}_k - \mathbf{x}) = \lambda \mathbf{y} + \frac{1}{k}(\mathbf{x}_0 - \mathbf{x})$,
      - and $\mathbf{x} + \mathbf{z}_k = \lambda \mathbf{y} + \frac{1}{k}\mathbf{x}_0 + (1 - \frac{1}{k})\mathbf{x} = \frac{1}{k}\mathbf{w}_k + (1 - \frac{1}{k})\mathbf{x} \in C$ by convex set.
      - By closeness of C, $\lim_{k \to \infty} \mathbf{x} + \mathbf{z}_k = \mathbf{x} + \mathbf{z} \in C$
- Definition (recession cone) The *recession cone* $R(C)$ of C is the set of directions of recession of C
  - This set is closed under non-negative linear combinations, see next lemma.
- Lemma 2.32 (non-negative linear combination of recession cone) Let $C \subseteq \mathbb{R}^d$ be closed convex set, and let $\mathbf{y_1}, \mathbf{y_2}$ be directions of recession of C. Then $\forall \lambda_1, \lambda_2 \in \mathbb{R}^+, \mathbf{y} = \lambda_1 \mathbf{y}_1 + \lambda_2 \mathbf{y}_2$ is also direction of recession of C.
  - Proof: $\mathbf{x} + \lambda \mathbf{y} = \mathbf{x} + \lambda_1 \lambda \mathbf{y}_1 + \lambda_2 \lambda \mathbf{y}_2 = \lambda_1 (\mathbf{x} + \lambda \mathbf{y}_1) + \lambda_2 (\mathbf{x} + \lambda \mathbf{y}_2) \in C$
- Definition 2.33 (direction of constancy) Let $C \subseteq \mathbb{R}^d$ be a convex set. Then $\mathbf{y} \in \mathbf{R}^d$ is a *direction of constancy* of C if both $\mathbf{y}$ and $-\mathbf{y}$ are directions of recession of C.
- Lemma 2.34 (linear combination). Let $C \subseteq \mathbb{R}^d$ be closed convex set, and let $\mathbf{y_1}, \mathbf{y_2}$ be directions of constancy of C. Then $\forall \lambda_1, \lambda_2 \in \mathbb{R}$, $\mathbf{y} = \lambda_1 \mathbf{y}_1 + \lambda_2 \mathbf{y}_2$ is also direction of constancy of C.

Recession Cone in Sub-level set

- Lemma (non-deceasing along direction) Suppose $\mathbf{y}$ is a direction of recession of $f^{\leq \alpha}$, then $\forall \mathbf{x} \in f^{\leq \alpha}, \forall \lambda \geq 0, f(\mathbf{x} + \lambda \mathbf{y}) \leq f(\mathbf{x})$
  - Proof
    - Fix λ and let $\mathbf{z} = \lambda \mathbf{y}$, define $\mathbf{w}_k := \mathbf{x} + k\mathbf{z}$, then $\mathbf{x} + \mathbf{z} = (1 - \frac{1}{k})\mathbf{x} + \frac{1}{k}\mathbf{w}_k$.
    - We know $f(\mathbf{w}_k) \leq \alpha$, then $f(\mathbf{x} + \mathbf{z}) \leq (1 - \frac{1}{k}) f(\mathbf{x}) + \frac{1}{k}\alpha \to f(\mathbf{x})$.
- Lemma 2.35 (level invariance of Recession cone) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Any two sublevel sets $f \leq \alpha, f \leq \alpha'$ have the same recession cones.
  - Proof Since two sets non-empty, $\exists \mathbf{x}' \in f^{\leq \alpha} \cap f^{\leq \alpha'}$, then $f(\mathbf{x}' + \lambda \mathbf{y}) \leq f(\mathbf{x}') \leq \min\{\alpha', \alpha\}$. Every direction of recession $\mathbf{y}$ of one set is also that of the other,
  - This Lemma gives the definition of cone of a function $R(f)$
- Definition 2.36 (recession cone / lineality space) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function.

- Then $\mathbf{y} \in \mathbb{R}^d$ is a direction of recession (of constancy, respectively) of f if $\mathbf{y}$ is a direction of recession (of constancy, respectively) for some (equivalently, for every) nonempty sublevel set.
- The set of directions of recession of f is called the *recession cone* $R(f)$ of f.
- The set of directions of constancy of f is called the *lineality space* $L(f)$ of f.

- Lemma 2.37 & 2.38 (Relation to Epigraph) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. The following statements are equivalent.
  - (i) $\mathbf{y} \in \mathbb{R}^d$ is a direction of (recession / constancy) of f.
  - (ii) $\forall \mathbf{x} \in \mathbb{R}^d, \lambda > 0$, ($f(\mathbf{x} + \lambda \mathbf{y}) \le f(\mathbf{x})$ / $f(\mathbf{x} + \lambda \mathbf{y}) = f(\mathbf{x})$)
  - (iii) $(\mathbf{y}, 0)$ is a direction of (recession / constancy) of (the closed convex set) $\mathbf{epi}(f)$.
  - Proof
    - (i)<->(ii) obvious
    - (ii)<->(iii) $(\mathbf{x}, f(\mathbf{x})) + \lambda(\mathbf{y}, 0) = (\mathbf{x} + \lambda \mathbf{y}, f(\mathbf{x}))$. $f(\mathbf{x} + \lambda \mathbf{y}) \le f(\mathbf{x}) \Leftrightarrow (\mathbf{x} + \lambda \mathbf{y}, f(\mathbf{x})) \in \mathbf{epi}(f)$.

Coercive convex functions (Boundedness of sublevel)

- Definition 2.39 (coerciveness) A convex function f is coercive if its recession cone is trivial, meaning that $\mathbf{0}$ is its only direction of recession.
  - Coercivity means that along any direction, $f(\mathbf{x})$ goes to infinity.
  - Coercive example: $x_1^2 + x_2^2$; non-coercive example: $f(x) = x, e^x$.
- Lemma 2.40 (Boundedness of coercive sublevel) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a coercive convex function. Then every nonempty sublevel set $f^{\le \alpha}$ is bounded.
  - Proof
    - Given sublevel set $f^{\le\alpha}$, for an $\mathbf{x} \in f^{\le\alpha}$, define mapping from $S^{d-1} = \{\mathbf{y} \in \mathbb{R}^d : \|y\| = 1\}$ to $\mathbb{R}$: $g(\mathbf{y}) = \max\{\lambda \ge 0 : f(\mathbf{x} + \lambda \mathbf{y}) \le \alpha\}$.
      - Since f coercive, g is well defined.
    - We claim g is continuous, by showing every sequence to $\mathbf{y}$ have function value converge to $g(\mathbf{y})$
      - For arbitary $\varepsilon > 0$, define $\{\underline{\lambda}, \overline{\lambda}\} := \{g(\mathbf{y}) - \varepsilon, g(\mathbf{y}) + \varepsilon\}$. so $f(\mathbf{x} + \underline{\lambda}\mathbf{y}) \le \alpha$ and $f(\mathbf{x} + \overline{\lambda}\mathbf{y}) > \alpha$
      - For every sequence $\{\mathbf{y}_k\}$ s.t. $\lim_{k\to\infty} \mathbf{y}_k = \mathbf{y}$, by the continuity of f, we have $\lim_{k\to\infty} f(\mathbf{x} + \underline{\lambda}\mathbf{y}_k) = f(\mathbf{x} + \underline{\lambda}\mathbf{y}) \le \alpha$ and $\lim_{k\to\infty} f(\mathbf{x} + \overline{\lambda}\mathbf{y}_k) = f(\mathbf{x} + \overline{\lambda}\mathbf{y}) > \alpha$
      - Therefore, since $\lim_{k\to\infty} f(\mathbf{x} + g(\mathbf{y}_k)\mathbf{y}_k) = \alpha$, for sufficiently large k, we have $\underline{\lambda} \le g(\mathbf{y}_k) \le \overline{\lambda}$.
      - This means continuity.
    - Since g continuous and S^{d-1} compact, g attains maximum of λ^*. For arbitary $\mathbf{x}' \in f^{\le\alpha}$, $\|\mathbf{x}' - \mathbf{x}\| \le g(\frac{\mathbf{x}' - \mathbf{x}}{\|\mathbf{x}' - \mathbf{x}\|}) \le \lambda^*$, Bounded.
- Theorem 2.41 (Global minimum). Let $f : \mathbb{R}^d \to \mathbb{R}$ be a coercive convex function. Then f has a global minimum.

Weakly coercive convex function

- Definition 2.42 Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Function f is called *weakly coercive* if its recession cone equals its lineality space.
  - example $f(x_1, x_2) = x_1^2$
- Theorem 2.43. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a weakly coercive convex function. Then f has a global minimum.
  - Proof Lineality space L is a linear subspace of $\mathbb{R}^d$, therefore, its orthogonal complement $L^\perp$ is coercive, and can obtain global minimum in $L^\perp$, which can be shown is also global minimum over all space by orthogonal decomposition.

Convex Programming

- Definition (Convex Program) minimize $f_0(\mathbf{x})$, subject to $f_i(\mathbf{x}) \le 0, i = 1, \ldots, m$ and $h_i(\mathbf{x}) = 0, i = 1, \ldots, p$, where f_i convex and h_i affine functions with domain $\mathbb{R}^d$.
  - Domain $\mathcal{D} = \left(\cap_{i=0}^m \operatorname{dom}(f_i)\right) \cap \left(\cap_{i=1}^p \operatorname{dom}(h_i)\right)$ is also convex.
  - $X = \{\mathbf{x} \in \mathbb{R}^d : f_i(\mathbf{x}) \le 0, i = 1, \ldots, m; h_i(\mathbf{x}) = 0, i = 1, \ldots, p\}$ is the *feasible region* of the program. $x \in X$ is called the *feasible solution*.

Lagrange duality (Weak duality)

- Idea: Turn hard constrains of primal into soft constrains into objective function.
- Definition (Lagrangian) givem convex program $\{f_i, h_i\}$, its *Lagrangian* $L : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(\mathbf{x})$.
- Definition (Lagrange dual function) The Lagrange dual function is the function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ defined by

$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$.
  - g assume value $-\infty$ is typical. The interesting $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are those $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty$.
- Lemma 2.45 (Weak Lagrange duality, lower bound of minimum) $\forall x \in X, \forall \boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^p$ s.t. $\lambda \geq 0$, we have $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\mathbf{x})$
  - Proof $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \underbrace{\sum_{i=1}^m \lambda_i f_i(\mathbf{x})}_{\leq 0} + \underbrace{\sum_{i=1}^p \nu_i h_i(\mathbf{x})}_{=0} \leq f_0(\mathbf{x})$
  - Finding the maximum g makes it the lower bound.
- Definition 2.46 (dual problem of original problem) Maximize $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ subject to $\lambda \geq 0$.
- Lemma (convexity of dual) Given definition of function $f : \mathbf{dom}(f) \to \mathbb{R} \cup \{\infty\}$ being convex if $\mathbf{dom}(f)$ and any finite $f(x), f(y) < \infty$ follow convex inequality. Then $-g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is convex.
  - Proof
    - Denote $f(\mathbf{x}, \mathbf{z}) := -L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}), f(\mathbf{z}) := -g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \sup_{x\in X} f(\mathbf{x}, \mathbf{z})$.
    - For $f(\mathbf{y}), f(\mathbf{z}) < \infty$, we have $f(\lambda\mathbf{y} + (1-\lambda)\mathbf{z}) = \sup_{\mathbf{x}\in X} f(\mathbf{x}, \lambda\mathbf{y} + (1-\lambda)\mathbf{z}) \leq \sup_{\mathbf{x}\in X} [\lambda f(\mathbf{x}, \mathbf{y}) + (1-\lambda) f(\mathbf{x}, \mathbf{z})]$ last by convexity
    - Since $f(\mathbf{y}), f(\mathbf{z}) < \infty$, $\sup_{x\in X} f(\mathbf{x}, \mathbf{z})$ and $\sup_{x\in X} f(\mathbf{x}, \mathbf{y})$ exists, So RHS $\leq \lambda \sup_{x\in X} f(\mathbf{x}, \mathbf{y}) + (1-\lambda) \sup_{x\in X} f(\mathbf{x}, \mathbf{z}) = \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{z})$

Strong duality

- Theorem 2.47 (Strong duality for convex program) Suppose a convex program has a feasible solution $\tilde{\mathbf{x}}$ that <u>in addition</u> satisfies $f_i(\tilde{\mathbf{x}}) < 0, i = 1, \ldots, m$ (*Slater point*). Then
  - (i) $\inf_{\mathbf{x}\in X} f_0(\mathbf{x}) = \sup_{\boldsymbol{\lambda}>0, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) := f^\star$, infimum of primal = supremum of dual.
  - (ii) if $|f^\star| < \infty$, then exists feasible solution of dual $\exists \boldsymbol{\lambda}^\star > 0, \boldsymbol{\nu}^\star$ s.t. $g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = f^\star$
  - *no proof*
- In practice, we minimize $f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(\mathbf{x})$ without constraint.
  - If Strong duality holds, $\exists \boldsymbol{\lambda}^\star > 0, \boldsymbol{\nu}^\star$ that have same infimum as unconstrained one.
  - Even if not, the infimum of unconstrained problem is a lower bound of primal.
- Strong duality may also hold when there is no Slater point, or even when not a convex program. Theorem 2.47 is only a sufficient condition.

Karush-Kuhn-Tucker (KKT) conditions

- *Key Idea*
  - If optimization program is differentiable, KKT condition is necessary
  - Futher if program is convex, then KKT condition is sufficient.
- Definition 2.48 (Zero duality gap) Let $\tilde{\mathbf{x}}$ be feasible for the primal and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ feasible for the Lagrange dual. The primal and dual solutions $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ are said to have *zero duality gap* if $f_0(\tilde{\mathbf{x}}) = g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$.
- Lemma 2.49 (Complementary slackness) If $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap, then $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, i = 1, \ldots, m$.
  - Proof
    - $f_0(\tilde{\mathbf{x}}) = g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = \inf_{\mathbf{x}\in\mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\mathbf{x})\right) \leq f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^m \underbrace{\tilde{\lambda}_i f_i(\tilde{\mathbf{x}})}_{\leq 0} + \sum_{i=1}^p \underbrace{\tilde{\nu}_i h_i(\tilde{\mathbf{x}})}_{0} \leq f_0(\tilde{\mathbf{x}})$
    - equation holds only when $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, i = 1, \ldots, m$.
- Lemma 2.50 (Vanishing Lagrangian gradient) If $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ have zero duality gap, and if all f_i and h_i are differentiable, then $\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = \mathbf{0}$
  - Proof Since $f_0(\tilde{\mathbf{x}}) = g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = \inf_{\mathbf{x}\in\mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\mathbf{x})\right)$, $\tilde{\mathbf{x}}$ is the minimizer of $L(\mathbf{x}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$
- Theorem 2.51 (=2.49 + 2.50, KKT necessary condition, zero gap -> KKT) Let $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ are feasible solution for primal and dual, and have zero duality gap, and if all f_i and h_i are differentiable, then
  - (i) $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, \ldots, m$
  - (ii) $\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = \mathbf{0}$
  - PS: no need to assume convexity
- Theorem 2.52 (KKT sufficient condition, convexity + KKT -> zero gap) Let $\tilde{\mathbf{x}}$ and $(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ are feasible solution for primal and dual, and all f_i and h_i are differentiable, all f_i are convex and all h_i affine. If KKT condition ((i) and (ii) in Theorem 2.51)holds, then $f_0(\tilde{\mathbf{x}}) = g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$ (zero duality gap).
  - Proof
    - By KKT(ii) $\nabla f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{\mathbf{x}}) = \mathbf{0}$, $\tilde{\mathbf{x}}$ is solution for unconstraint convex optimization problem $\min_{\mathbf{x}\in\mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\mathbf{x})\right)$ (Lemma 2.21), therefore $g(\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}})$.

- By KKT(i) $\tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) = 0, \quad i = 1, \ldots, m$ and because $\tilde{\mathbf{x}}$ feasible,
  $L(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\nu}}) = f_0(\tilde{\mathbf{x}}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{\mathbf{x}}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{\mathbf{x}}) = f_0(\tilde{\mathbf{x}})$
- We get duality gap.