MA3236 Non-Linear Programming

AY2022/23 Semester 1 · *Prepared by Tian Xiao @snoidetx*

1 Optimization Problem

Bounded set: $\exists M > 0\, \forall \boldsymbol{x} \in S\, [||\boldsymbol{x}|| \leq M]$.
Compact set: closed & bounded.
Local minimizer: $\exists \varepsilon > 0\, [f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)\, \forall x \in S \cap B(\boldsymbol{x}^*, \varepsilon)]$.
Global minimizer: $\forall \boldsymbol{x} \in S\, [f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)]$.
Weierstrass Theorem A continuous function on a non-empty compact set $S \subset \mathbb{R}^n$ has global maximizer and minimizer.

2 Convex Optimization

Convex set: $\boldsymbol{x}, \boldsymbol{y} \in D \Rightarrow \forall \lambda \in [0,1]\, [\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in D]$.
- Intersection of convex sets is convex. Union may not be convex.

Convex function: Let $D \subseteq \mathbb{R}^n$ be a convex set. $f : D \to \mathbb{R}$ is a convex/concave function if $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq / \geq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$.
- Suppose f_1 and f_2 are convex:
  (a) $f_1 + f_2$ is convex; (b) $\max(f_1, f_2)$ is convex.
  (c) αf_1 is convex for $\alpha \geq 0$, concave for $\alpha < 0$;
- f_j is convex $\Rightarrow f(\boldsymbol{x}) = \sum_{j=1}^{k} \alpha_j f_j(\boldsymbol{x}), \alpha \geq 0$ is convex.
- h convex, g non-de(in)creasing convex $\Rightarrow g \circ h$ convex(concave).
- D, f convex $\Rightarrow \forall \alpha \in \mathbb{R}\, [S_\alpha = \{\boldsymbol{x} \in D : f(\boldsymbol{x}) \leq \alpha \text{ is convex}\}]$.
- f convex, $\boldsymbol{x} = \sum_{j=1}^{k} \lambda_j \boldsymbol{x}^{(j)}, \sum_{j=1}^{k} \lambda_j = 1 \Rightarrow f(\boldsymbol{x}) \leq \sum_{j=1}^{k} \lambda_j f(\boldsymbol{x}^{(j)})$.

Gradient vector: $\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \cdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{pmatrix}$.
- $\nabla f(\boldsymbol{x}^*)^\top \mathbf{d} = \lim_{\lambda \to 0} \frac{f(\boldsymbol{x}^* + \lambda \mathbf{d}) - f(\boldsymbol{x}^*)}{\lambda}$.
- At $\boldsymbol{x}^*$, $f(\boldsymbol{x})$ decreases most rapidly along the direction $-\nabla f(\boldsymbol{x}^*)$ and increases most rapidly along the direction $\nabla f(\boldsymbol{x}^*)$.

Tangent Plane Characterization
(a) f is convex. $\Leftrightarrow f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \leq f(\boldsymbol{y}), \forall \boldsymbol{x}, \boldsymbol{y} \in S$.
(b) f is strictly convex. $\Leftrightarrow f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) < f(\boldsymbol{y}), \forall \boldsymbol{x} \neq \boldsymbol{y} \in S$.
Theorem 4.9. $\boldsymbol{x}^*$ global minimizer $\Leftrightarrow \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \forall x \in C$.

Hessian: $H_f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{x}) \end{pmatrix}$.

Second Order Test Suppose f has continuous second order derivative:

+ve semidefinite	$\forall \boldsymbol{x} \in \mathbb{R}^n\, [\boldsymbol{x}^\top A \boldsymbol{x} \geq 0]$	$\forall \lambda\, [\lambda \geq 0]$	$\Leftrightarrow$ convex
+ve definite	$\forall \boldsymbol{x} \neq \mathbf{0}\, [\boldsymbol{x}^\top A \boldsymbol{x} > 0]$	$\forall \lambda\, [\lambda > 0]$	$\Rightarrow$ strictly convex
-ve semidefinite	$\forall \boldsymbol{x} \in \mathbb{R}^n\, [\boldsymbol{x}^\top A \boldsymbol{x} \leq 0]$	$\forall \lambda\, [\lambda \leq 0]$	$\Leftrightarrow$ concave
-ve definite	$\forall \boldsymbol{x} \neq \mathbf{0}\, [\boldsymbol{x}^\top A \boldsymbol{x} < 0]$	$\forall \lambda\, [\lambda < 0]$	$\Rightarrow$ strictly concave
indefinite	none of the above	$\begin{cases} \lambda_1 > 0 \\ \lambda_2 < 0 \end{cases}$	$\Rightarrow$ neither nor

Principal minor: $\Delta_k = \det \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}$.
- $\forall k\, [\Delta_k > 0] \Rightarrow$ +ve definite; $\forall k\, [(-1)^k \Delta_k > 0] \Rightarrow$ -ve definite.

Taylor Theorem If f has continuous second order derivative, then $\exists w \in [x, y]\, [f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}(y - x)^\top H_f(w)(y - x)]$.

3 Unconstrained Optimization

Coercive function: $\lim_{||\boldsymbol{x}|| \to \infty} f(\boldsymbol{x}) = +\infty$.
Theorem 6.4. continuous & coercive $\Rightarrow$ at least 1 global minimizer.
Stationary point: $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.
- If f has continuous 1st and 2nd order derivative, then $\boldsymbol{x}^*$ is a local minimizer $\Rightarrow \nabla f(\boldsymbol{x}^*) = \mathbf{0} \Rightarrow H_f(\boldsymbol{x}^*)$ is +ve semidefinite.

Saddle point: stationary & not local minimizer/maximizer.
- stationary & $H_f(\boldsymbol{x}^*)$ is indefinite $\Rightarrow \boldsymbol{x}^*$ is saddle point.

Theorem 7.7. stationary & +ve/-ve definite $\Rightarrow \boldsymbol{x}^*$ strict local optimal.
Theorem 7.10. (strictly) convex & local min $\Rightarrow$ (unique) global min.
Corollary 7.11. f is convex & $\boldsymbol{x}^*$ stationary $\Rightarrow \boldsymbol{x}^*$ is global minimizer.
Theorem 7.15. $\boldsymbol{x}^*$ is a global minimizer of q and $\boldsymbol{Q}$ is symmetric +ve semidefinite. Then $q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} \Leftrightarrow \boldsymbol{Q}\boldsymbol{x}^* = -\boldsymbol{c}$.

3.1. Bisection Method

Intermediate Value Theorem $f'(a)f'(b) < 0 \Rightarrow \exists r \in (a, b)\, [f(r) = 0]$.
Bisection Search Algorithm Set tolerance $\varepsilon > 0$.
[Step 1] Choose $[a_1, b_1]$ such that $f'(a_1)f'(b_1) < 0$.
[Step k] Set $x_k = \frac{1}{2}(a_k + b_k)$. If $b_k - a_k \leq 2\varepsilon$, return x_k; else, if $f'(a_k)f'(x_k) < 0$, set $[a_{k+1}, b_{k+1}] = [a_k, x_k]$, vice versa.

3.2. One-Variable Newton's Method

Taylor's approx: $f(x) \approx f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2$.
Newton's Method Algorithm Set initial point x_0 and tolerance ε.
[Step k] If $|f'(x_k)| < \epsilon$, return x_k; else, compute $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$.

3.3. Golden Section Method

Unimodal function: f is unimodal on $[a, b]$ if it has exactly one global minimizer x^* on $[a, b]$ and it is strictly decreasing on $[a, x^*]$ and strictly increasing on $[x^*, b]$.
Golden Section Method Algorithm
[Step 0] Set $[a_0, b_0] = [a, b]$, $\varepsilon > 0$, $\alpha = \frac{\sqrt{5}-1}{2}$. Compute $\lambda_0 = b - \alpha(b - a)$, $\mu_0 = a + \alpha(b - a)$, $f(\lambda_0)$ and $f(\mu_0)$.
[Step k] If $b_k - a_k < \varepsilon$, return $x^* \in [a_k, b_k]$; else, if $f(\lambda_k) > f(\mu_k)$, set $a_{k+1} = \lambda_k$, $b_{k+1} = b_k$, $\lambda_{k+1} = \mu_k$, $\mu_{k+1} = \lambda_k + \alpha(b_k - \lambda_k)$, vice versa.

General Optimization Algorithm Framework

$\boldsymbol{x}^{(0)} \leftarrow$ some initial guess
for $k = 0, 1, \cdots$ **do**
  if $\boldsymbol{x}^{(k)}$ is optimal **then return** $\boldsymbol{x}^{(k)}$
  else
    $\boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}$
  end if
end for

3.4. Multi-Variable Newton Method

Newton's Method Algorithm Set initial point $\boldsymbol{x}^{(0)}$ and tolerance ε.
[Step k] If $||\nabla f(\boldsymbol{x}^{(k)})|| < \varepsilon$, return $\boldsymbol{x}^{(k)}$; else, compute $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{p}^{(k)}$, where $\boldsymbol{p}^{(k)} = -H_f\left(\boldsymbol{x}^{(k)}\right)^{-1} \nabla f(\boldsymbol{x}^{(k)})$.
- Pro: Convergent is fast (quadratically near the solution), assuming $H_f(\boldsymbol{x}^*)$ is non-singular and Lipschitz continuous in a neighbourhood of $\boldsymbol{x}^*$.
- Con: Computational cost per iteration is expensive.

Exact line search: Choose $\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)})$.
Armijo line search: Set $\sigma \in (0, 0.5)$ and $\beta \in (0, 1)$. Recursive set $\alpha \leftarrow \beta\alpha$ until $f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}) \leq f(\boldsymbol{x}^{(k)}) + \alpha\sigma \nabla f(\boldsymbol{x}^{(k)})^\top \boldsymbol{p}^{(k)}$.

3.5. Steepest Descent Method

Steepest Descent Algorithm
[Step 0] Set initial guess $\boldsymbol{x}^{(0)}$, tolerance ε.
[Step k] If $\boldsymbol{p}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) < \varepsilon$, stop.
Zig-zag behaviour: Moves in perpendicular steps (slow and inefficient).
Properties of steepest descent:
(a) Monotonic decreasing.
(b) If f is coercive, any convergent subsequence of $\boldsymbol{x}^{(k)}$ from steepest descent methods converge to a critical point of f.

3.6. Conjugate Gradient Method

Conjugate vector: $(\boldsymbol{p}^{(i)})^\top A \boldsymbol{p}^{(j)} = \mathbf{0}$.
Convex quadratic program (CQP): $\min_{\boldsymbol{x} \in \mathbb{R}^n} \phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}$.
Conjugate Gradient Algorithm Let $\boldsymbol{r}^{(0)} \leftarrow \boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b}$, $\boldsymbol{p}^{(0)} \leftarrow -\boldsymbol{r}^{(0)}$.
[Step k] If $\boldsymbol{r}^{(k)} = 0$, return $\boldsymbol{x}^{(k)}$; else, $\alpha_k \leftarrow -\frac{(\boldsymbol{r}^{(k)})^\top \boldsymbol{p}^{(k)}}{(\boldsymbol{p}^{(k)})^\top \boldsymbol{A} \boldsymbol{p}^{(k)}}$, $\boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{p}^{(k)}$, $\boldsymbol{r}^{(k+1)} \leftarrow \boldsymbol{A}\boldsymbol{x}^{(k+1)} - \boldsymbol{b} = \boldsymbol{r}^{(k)} + \alpha_k \boldsymbol{A}\boldsymbol{p}^{(k)}$, $\beta_{k+1} = \frac{(\boldsymbol{r}^{(k+1)})^\top \boldsymbol{A} \boldsymbol{p}^{(k)}}{(\boldsymbol{p}^{(k)})^\top \boldsymbol{A} \boldsymbol{p}^{(k)}}$, $\boldsymbol{p}^{(k+1)} = -\boldsymbol{r}^{(k+1)} + \beta_{k+1}\boldsymbol{p}^{(k)}$.

Method for CQP	Time	Convergence
Direct method: $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$	$O(n^3)$	N.A.
Steepest descent with exact line search	$O(n^2)$	slow (linear)
Newton's method	$O(n^3)$	1
Conjugate gradient	$O(n^2)$	$\leq n$

4 Constrained Optimization

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \quad | \quad \boldsymbol{x} \in \mathbb{R}^n \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \cdots, m \\ & h_j(\boldsymbol{x}) \leq 0, \quad j = 1, 2, \cdots, p \end{aligned}$$

Active constraint: $h_j(\boldsymbol{x}^*) = 0 \Rightarrow h_j$ is active at $\boldsymbol{x}^*$.
Regular point: $\boldsymbol{x}^*$ is a feasible point. Let $J(\boldsymbol{x}^*) = \{j \in \{1, 2, \cdots, p\} : h_j(\boldsymbol{x}^*) = 0\}$ be the index set of active constraints at $\boldsymbol{x}^*$. If the set $\{\nabla g_i(\boldsymbol{x}^*) : i = 1, 2, \cdots, m\} \cup \{\nabla h_j(\boldsymbol{x}^*) : j \in J(\boldsymbol{x}^*)\}$ is linearly independent, then $\boldsymbol{x}^*$ is a regular point.

KKT 1^{st} Order Condition $\boldsymbol{x}^*$ is a KKT point if it is a regular point and satisfies KKT first order necessary condition $\exists\lambda_1^*,\cdots,\lambda_m^*,\mu_1^*,\cdots,\mu_p^*\left[\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\lambda_i^*\nabla g_i(\boldsymbol{x}^*)+\sum_{j=1}^{p}\mu_j^*\nabla h_j(\boldsymbol{x}^*)=0\right]$, where $\forall j=1,2,\cdots,p\,[\mu_j^*\geq 0]$ and $\forall j\notin J(\boldsymbol{x}^*)\,[\mu_j^*=0]$.
Lagrange multiplier: $\lambda_i, i=1,2,\cdots,m$ and $\mu_j, j=1,2,\cdots,p$.
Complementary Slack Condition $\forall j=1,2,\cdots,p\,[\mu_j^*h_j(\boldsymbol{x}^*)=0]$.
KKT 2^{nd} Order Condition
$\boldsymbol{x}^*$ is a KKT point & $\forall \boldsymbol{y}\in C(\boldsymbol{x}^*,\lambda^*,\mu^*)\,[\boldsymbol{y}^\top H_L(\boldsymbol{x}^*)\boldsymbol{y}\geq 0]$.
Here $H_L(\boldsymbol{x}^*)=H_f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\lambda_i^*H_{g_i}(\boldsymbol{x}^*)+\sum_{j=1}^{p}\mu_j^*H_{h_j}(\boldsymbol{x}^*)$;

$$C(\boldsymbol{x}^*,\lambda^*,\mu^*)=\left\{\boldsymbol{y}\in\mathbb{R}^n:\begin{cases}\nabla g_i(\boldsymbol{x}^*)^\top\boldsymbol{y}=0 & i=1,2,\cdots,m\\ \nabla h_j(\boldsymbol{x}^*)^\top\boldsymbol{y}=0 & j\in J(\boldsymbol{x}^*)\text{ and }\mu_j>0\\ \nabla h_j(\boldsymbol{x}^*)^\top\boldsymbol{y}\leq 0 & j\in J(\boldsymbol{x}^*)\text{ and }\mu_j=0\end{cases}\right\}.$$

Strict complementarity: $\forall j\in J(\boldsymbol{x}^*)\,[\mu_j>0]$.
Theorem 13.3. $\boldsymbol{x}^*$ is regular & local minimizer $\Rightarrow$ 1^{st} & 2^{nd} condition.
Corollary 13.4. $\boldsymbol{x}^*$ is global minimizer $\Rightarrow$ $\boldsymbol{x}^*$ is KKT point.
Proposition 13.1. Suppose $\boldsymbol{x}^*\in S$ and strict complementarity holds at $\boldsymbol{x}^*$, then $\forall\boldsymbol{y}\in C(\boldsymbol{x}^*,\lambda^*,\mu^*)\,[\boldsymbol{y}^\top H_L(\boldsymbol{x}^*)\boldsymbol{y}\geq 0]\Leftrightarrow Z(\boldsymbol{x}^*)^\top H_L(\boldsymbol{x}^*)Z(\boldsymbol{x}^*)$ is +ve semidefinite. $Z(\boldsymbol{x}^*)\in\mathbb{R}^{n\times q}$ is a matrix whose columns form a basis of null space of $(\nabla g_1(\boldsymbol{x}^*),\cdots,\nabla g_m(\boldsymbol{x}^*),[\nabla h_j(\boldsymbol{x}^*):j\in J(\boldsymbol{x}^*)])^\top$.
Theorem 13.8. KKT point & $H_L(\boldsymbol{x}^*)$ +ve definite on $C(\boldsymbol{x}^*,\lambda^*,\mu^*)\Rightarrow$ $\boldsymbol{x}^*$ is strict local minimizer.
Theorem 14.1. f and h_j are differentiable convex functions & $g_i(\boldsymbol{x})=\boldsymbol{A}_i^\top\boldsymbol{x}-b_i\implies$ (KKT $\Rightarrow$ global minimizer).
Slater's condition: $\exists\hat{\boldsymbol{x}}\in\mathbb{R}^n\,[\forall i=1,2,\cdots,m\,[g_i(\hat{\boldsymbol{x}})=0]\,\&\,\forall j=1,2,\cdots,p\,[h_j(\hat{\boldsymbol{x}})<0]]$.
Theorem 14.4. f and h_j are differentiable convex functions & $g_i(\boldsymbol{x})=\boldsymbol{A}_i^\top\boldsymbol{x}-b_i$ & $p\geq 1$ & Slater's condition $\implies$ (global minimizer $\Rightarrow$ KKT).
Linear equality constrained convex program (ECP):

$$\begin{aligned}\min\quad & f(\boldsymbol{x})\quad|\quad\boldsymbol{x}\in\mathbb{R}^n\\ \text{s.t.}\quad & \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\end{aligned}$$

Where rows of $\boldsymbol{A}$ are linearly independent; f is differentiable & convex.
Theorem 14.6. In an ECP, KKT $\Leftrightarrow$ global minimizer. In an ENLP (ECP without convexity), global minimizer $\Rightarrow$ KKT.
Primal problem and Lagrangian dual problem:

$$\begin{aligned}\min\quad & f(\boldsymbol{x})\\ \text{s.t.}\quad & g_i(\boldsymbol{x})=0,\quad i=1,2,\cdots,m\\ & h_j(\boldsymbol{x})\leq 0,\quad j=1,2,\cdots,p\\ & \boldsymbol{x}\in X\subseteq\mathbb{R}^n\end{aligned}\iff\begin{aligned}&\max_{\boldsymbol{\lambda}\in\mathbb{R}^m,\boldsymbol{\mu}\in\mathbb{R}^p_+}\theta(\boldsymbol{\lambda},\boldsymbol{\mu})\\ &\theta(\boldsymbol{\lambda},\boldsymbol{\mu})=\inf_{\boldsymbol{x}\in X}\{f(\boldsymbol{x})+\boldsymbol{\lambda}^\top\mathbf{g}(\boldsymbol{x})\\ &\qquad+\boldsymbol{\mu}^\top\mathbf{h}(\boldsymbol{x})\}\end{aligned}$$

- θ is finite $\Rightarrow$ θ is concave.

Weak Duality Theorem Let $\boldsymbol{x}$ be an optimal solution to (P) and $(\boldsymbol{\lambda},\boldsymbol{\mu})$ be an optimal solution to (D). Then $f(\boldsymbol{x})\geq\theta(\boldsymbol{\lambda},\boldsymbol{\mu})$.
Strong Duality Theorem X convex & f and h_j convex & g_i affine & Slater's condition $\Rightarrow$ duality gap is zero (i.e. $\inf f(\boldsymbol{x})=\sup\theta(\boldsymbol{\lambda},\boldsymbol{\mu})$). If inf is attained at $\boldsymbol{x}^*$, then $\boldsymbol{\mu}^*\boldsymbol{h}(\boldsymbol{x}^*)=0$.
Saddle point: A point $(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ is a saddle point of $L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})=f(\boldsymbol{x})+\boldsymbol{\lambda}^\top\mathbf{g}(\boldsymbol{x})+\boldsymbol{\mu}^\top\mathbf{h}(\boldsymbol{x})$ if $\boldsymbol{x}^*\in X$ & $\boldsymbol{\mu}^*\geq 0$ & $L(\boldsymbol{x}^*,\boldsymbol{\lambda},\boldsymbol{\mu})\leq L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\leq L(\boldsymbol{x},\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)$ for all $\boldsymbol{x}\in X$ and all $(\boldsymbol{\lambda},\boldsymbol{\mu})$ with $\boldsymbol{\mu}\geq 0$.
Theorem 18.3. saddle point $\Rightarrow$ optimal solutions for (P) and (D).
Corollary 18.4. Saddle points are KKT points.
Theorem 18.5. KKT & f convex $\Rightarrow$ saddle point.
Corollary 18.6. KKT & f convex $\Rightarrow$ optimal solution of (P) and (D).

4.1. Subgradient Method

$$(D')\qquad\max_{\boldsymbol{w}\in\mathbb{R}^m\times\mathbb{R}^p_+}\theta(\boldsymbol{w})=\inf_{\boldsymbol{x}\in X}\left\{f(\boldsymbol{x})+\boldsymbol{w}^\top\boldsymbol{\beta}(\boldsymbol{x})\right\}$$

- $\boldsymbol{X}(\boldsymbol{w})$: The set of minimizers given $\boldsymbol{w}$.

Lemma 19.1. If $\boldsymbol{X}(\bar{\boldsymbol{w}})$ is singleton $\{\bar{\boldsymbol{x}}\}$, then for any $\boldsymbol{w}^{(k)}\to\bar{\boldsymbol{w}}$ and $\boldsymbol{x}^{(k)}\in\boldsymbol{X}^{(k)}$, we have $\boldsymbol{x}^{(k)}\to\bar{\boldsymbol{x}}$.
Theorem 19.2. If $\boldsymbol{X}(\bar{\boldsymbol{w}})$ is singleton $\{\bar{\boldsymbol{x}}\}$, then θ is differentiable at $\bar{\boldsymbol{w}}$, with gradient $\nabla\theta(\bar{\boldsymbol{w}})=\boldsymbol{\beta}(\bar{\boldsymbol{x}})$.
Subgradient $\boldsymbol{\xi}$: S,f convex, $\forall\boldsymbol{x}\in S\,[f(\boldsymbol{x})\geq f(\bar{\boldsymbol{x}})+\boldsymbol{\xi}^\top(\boldsymbol{x}-\bar{\boldsymbol{x}})]$.
Subdifferential $\delta f(\bar{\boldsymbol{x}})$: $\{\boldsymbol{\xi}:\boldsymbol{\xi}\text{ is subgradient}\}$.
Lemma 19.5. For any $\bar{\boldsymbol{x}}\in\boldsymbol{X}(\bar{\boldsymbol{w}})$, $\boldsymbol{\beta}(\bar{\boldsymbol{x}})$ is a subgradient of θ at $\bar{\boldsymbol{w}}$.
Directional derivative: $\varphi'(\bar{\boldsymbol{x}},\boldsymbol{d})=\lim_{\lambda\to 0^+}\frac{\varphi(\bar{\boldsymbol{x}}+\lambda\boldsymbol{d})-\varphi(\bar{\boldsymbol{x}})}{\lambda}$.
Theorem 19.7. If φ is concave at $\bar{\boldsymbol{x}}$, then $\varphi'(\bar{\boldsymbol{x}},\boldsymbol{d})$ exists.
Lemma 19.8. $\exists\bar{\boldsymbol{x}}\in\boldsymbol{X}\,[\theta'(\bar{\boldsymbol{x}},\boldsymbol{d})\geq\boldsymbol{d}^\top\boldsymbol{\beta}(\bar{\boldsymbol{x}})]$.
Theorem 19.9. $\theta'(\bar{\boldsymbol{x}},\boldsymbol{d})=\inf\{\boldsymbol{d}^\top\boldsymbol{\xi}:\boldsymbol{\xi}\in\delta\theta(\bar{\boldsymbol{w}})\}$.
Theorem 19.10. $\delta\theta(\bar{\boldsymbol{w}})=conv\{\boldsymbol{\beta}(\boldsymbol{y}):\boldsymbol{y}\in\boldsymbol{X}(\bar{\boldsymbol{w}})\}$.
Ascent direction $\boldsymbol{d}$: $\exists\delta>0\,\forall\lambda\in(0,\delta)\,[\theta(\boldsymbol{w}+\lambda\boldsymbol{d})>\theta\boldsymbol{w}]$.
Steepest ascent direction $\bar{\boldsymbol{d}}$: $\theta'(\boldsymbol{w},\bar{\boldsymbol{d}})=\max_{||\boldsymbol{d}||\leq 1}\theta'(\boldsymbol{w},\boldsymbol{d})$.
Theorem 20.9. Let $\hat{\boldsymbol{\xi}}$ be a subgradient in $\delta\theta(\boldsymbol{w})$ with the smallest Euclidean norm. Then $\bar{\boldsymbol{d}}=\begin{cases}\mathbf{0} & \hat{\boldsymbol{\xi}}=\mathbf{0}\\ \frac{\hat{\boldsymbol{\xi}}}{||\hat{\boldsymbol{\xi}}||} & \hat{\boldsymbol{\xi}}\neq\mathbf{0}\end{cases}$ is a direction of steepest ascent.

4.2. Frank-Wolfe Algorithm

ICP: $\min f(\boldsymbol{x})$ s.t. $\boldsymbol{A}\boldsymbol{x}\leq\boldsymbol{b}$. f is convex.
Frank-Wolfe Algorithm Set tolerance ε, $\boldsymbol{x}^{(0)}$. Set $\text{LB}_0\leftarrow-\infty$.
[Step k] Set $\text{UB}_k\leftarrow f(\boldsymbol{x}^{(k)})$. If $\text{UB}_k-\text{LB}_k\leq\varepsilon$, return $\boldsymbol{x}^{(k)}$; else, solve the LP_k : $\min z(\boldsymbol{x})=f(\boldsymbol{x}^{(k)})+\nabla f(\boldsymbol{x}^{(k)})^\top(\boldsymbol{x}-\boldsymbol{x}^{(k)})$ s.t. $\boldsymbol{A}\boldsymbol{x}\leq\boldsymbol{b}$. Let $\hat{\boldsymbol{x}}^{(k)}$ be the optimal solution and $\hat{z}_k$ be the optimal value. $\text{LB}_{k+1}\leftarrow\max(\text{LB}_k,\hat{z}_k)$. $\boldsymbol{p}^{(k)}\leftarrow\hat{\boldsymbol{x}}^{(k)}-\boldsymbol{x}^{(k)}$. Do line search to find optimal α_k. $\boldsymbol{x}^{(k+1)}\leftarrow\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{p}^{(k)}$.
- $\text{LB}=\hat{z}\leq f(\boldsymbol{x}^*)\leq f(\hat{\boldsymbol{x}})=\text{UB}$.

4.3. Quadratic Penalty Method

Equality constraint program (ECP): $\min f(\boldsymbol{x})$ s.t. $c_i(\boldsymbol{x})=0, i\in\mathcal{E}$.
Quadratic penalty function: $Q(\boldsymbol{x},\mu)=f(\boldsymbol{x})+\frac{1}{2\mu}\sum_{i\in\mathcal{E}}c_i^2(\boldsymbol{x})$.

Quadratic Penalty Algorithm Set tolerance ε, $\boldsymbol{x}^{(0)}$. $\mu_0\leftarrow 1$.
[Step k] Find an approximate minimizer of $Q(\boldsymbol{x},\mu_k)$ using Newton's method. If $||c(\boldsymbol{x}^{(k+1)})||\leq\varepsilon$, return $\boldsymbol{x}^{(k+1)}$; else, $\mu_{k+1}=\rho\mu_k,\rho<1$.
KKT condition: $||\nabla_{\boldsymbol{x}}Q(\boldsymbol{x}^{(k+1)},\mu_k)||\leq\tau_k$.

4.4. Augmented Lagrangian Method

AL: $L_A(\boldsymbol{x},\boldsymbol{\lambda},\mu)=f(\boldsymbol{x})+\sum_{i\in\mathbb{E}}\lambda_ic_i(\boldsymbol{x})+\frac{1}{2\mu}\sum c_i(\boldsymbol{x})^2$.

AL Algorithm Given $\mu_0>0,\tau_0>0$. Choose $\boldsymbol{x}^{(0)},\boldsymbol{\lambda}^{(0)}$.
[Step k] Find an approximate minimizer $\boldsymbol{x}^{(k+1)}$ of $L_A(\cdot,\boldsymbol{\lambda}^{(k)},\mu_k)$. If final convergence test is satisfied, return $\boldsymbol{x}^{(k+1)}$; else, $\lambda_i^{(k+1)}\leftarrow\lambda_i^{(k)}+\frac{c_i\boldsymbol{x}^{(k+1)}}{\mu_k}$. Choose new μ_{k+1},τ_{k+1}.

4.5. Barrier Function Method

Barrier function: $B(\boldsymbol{x})=\sum_{i\in I}\phi(-c_i(\boldsymbol{x}))$, where $\phi:\mathbb{R}^+\to\mathbb{R}^+$ s.t. $\phi'(x)<0$ and $\lim_{x\to 0^+}\phi(x)=\infty$ (e.g. $-\log(\cdot)$). $P(\boldsymbol{x},\mu_k)=f(\boldsymbol{x})+\mu_kB(\boldsymbol{x})$.
Barrier Function Algorithm Given $\mu_0>0,\tau_0>0$. Choose $\boldsymbol{x}^{(0)}$.
[Step k] Find an approximate minimizer $\boldsymbol{x}^{(k+1)}$ of $P(\boldsymbol{x},\mu_k)$. If final convergence test is satisfied, return $\boldsymbol{x}^{(k+1)}$; else, choose new $\mu_{k+1}\in(0,\mu_k),\tau_{k+1}$.

---

5 Summary