C0: Basic Topology (MA2108S Recap)

Rudin's Point Terminologies ($E^\circ =$ **Interior**(E), $\partial E =$ **Boundary**(E), $\bar{E} =$ **Closure**$(E) \overset{\text{def}}{=} E^\circ \oplus \partial E \overset{\text{def}}{=} E + E'$)

Limit Point $p \in E' : \forall r > 0(N_r(p) \backslash \{p\} = \emptyset)$ ($\because$deleted neighbourhood)
Isolated Point: $p \in E - E'$. $\therefore p \in E \Leftrightarrow p$ limit point xor isolated point.
Interior point $p \in E^\circ : \exists r > 0(N_r(p) \subseteq E \Leftrightarrow N_r(p) \cap E^c = \emptyset)$

Boundary point $p \in \partial E : \forall r > 0(N_r(p) \cap E \neq \emptyset \wedge N_r(p) \cap E^c \neq \emptyset) \overset{\text{def}}{\Leftrightarrow} p \in \bar{E} \cap \overline{E^c} \overset{\text{def}}{\Leftrightarrow} p \in \bar{E} - E^\circ$. $\therefore p \in \partial E \not\Rightarrow p \in E$.
Exterior Point: $\exists r > 0(N_r(p) \cap E = \emptyset \Leftrightarrow N_r(p) \subseteq E^c)$. $\therefore p \in E \Leftrightarrow p$ interior, boundary, xor exterior point.
Corollaries: 1. Limit point is interior xor boundary, and provably not exterior.
Purely-interior eg: $p = 0 \in \mathbb{R}$. Purely-boundary eg: $p = 0 \in \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}^+\}$.
2. Isolated point is interior xor boundary.
Purely-interior eg: $p = 0 \in \mathbb{Z}$. Purely-boundary eg: $p = 0 \in \{0\}$
3. If metric space X connected, then isolated $\Rightarrow$ boundary $\wedge E^\circ \subseteq E' \wedge (E' \neq \emptyset \Rightarrow E$ infinite)
4. $\forall p \in E'(\forall r > 0)(N_r(p) \cap E$ infinite)

Rudin's Set Terminologies [LUB/GCB: $\exists \min\{E\}, \max\{E\}$ of bdd E in $\mathbb{R}^k$]

Open set: $E = E^\circ$ | Closed set: $E' \subseteq E \overset{\text{def}}{\Leftrightarrow} E = \bar{E}$ | Crowded set: $E' \supseteq E$ (eg: $\mathbb{Q}^c$) | Perfect (closed crowded) set: $E' = E$
E bounded in $X : \exists r > 0(\exists \vec{q} \in X)(N_r(\vec{q}) \supseteq E) \Leftrightarrow \exists r > 0(\forall \vec{q} \in E)(\|\vec{q}\| \leq r)$ |
Corollaries: 1. de-Morgan's Law: $\forall A \succeq \mathbb{N}(\forall a \in A)((\bigcup_{a \in A} E_a)^c = \bigcap_{a \in A} E_a^c)$. $\therefore E$ open $\Leftrightarrow E$ closed.
2. Any set E is open and perfect in itself. $\bar{E} =$ min closed superset, $E^\circ =$ max open subset.
3. (Countable $\bigcup$/Finite $\bigcap$) keeps openness. (Countable $\bigcap$/Finite $\bigcup$) keeps closure
Countable open $\bigcap$ eg: $\bigcap_{n=1}^{\infty}(-\frac{1}{n}, \frac{1}{n}) = \{0\}$|Countable closed $\bigcup$ eg: $\bigcup_{n=1}^{\infty}[\frac{1-n}{n}, \frac{n-1}{n}] = (-1, 1)$

Relative Openness/Relative Closeness

$\forall E \subseteq$subspace $Y \subseteq$metric space $X(E$ open/closed in $Y \Leftrightarrow \exists$ open/closed $G \subseteq X(E = Y \cap G))$
E open in $X \Rightarrow E$ open in $Y \subseteq X$. E closed in $Y \Rightarrow E$ closed in $X \supseteq Y$.

Compactness

E compact in X: $\exists$ open $U_i \subseteq X \Rightarrow \exists n \in \mathbb{N}^+(\bigcup_{j=1}^{n} U_{i_j} \supseteq E)$ i.e, $\exists$ finite sub-cover per open cover.
$\forall E \subseteq$ subspace $Y \subseteq$ metric space $X(E$ compact in $Y \Leftrightarrow E$ compact in $X)$|$\forall$ closed $F \subseteq$ compact $E(F$ compact in $X)$
k-cell compact in $\mathbb{R}^k$|$\forall$ infinite $G \subseteq$ compact $E(E \cap G' \neq \emptyset)$
E compact $\Rightarrow E$ closed bdd in X. Converse eg: E closed bdd in $X = \mathbb{Q} \cap [0, 1]$ not complete ($\because \exists\{q_n\} \to \frac{\sqrt{2}}{2}$ divergent)

Heine-Borel Theorem: If $X = \mathbb{R}^k$, E compact $\Leftrightarrow E$ closed bounded

Perfectness

$\forall$ perfect $E \subseteq \mathbb{R}^k(E = \emptyset \vee E \succ \mathbb{N})$ | $\exists$ perfect, disconnected E eg: Cantor's set $= \bigcap_{n=1}^{\infty} \bigcup_{k=0}^{3^{n-1}-1}([\frac{3k}{3^n}, \frac{3k+1}{3^n}] \cup [\frac{3k+2}{3^n}, \frac{3k+3}{3^n}])$

Connectedness [$\forall A, B \subseteq X(A, B$ **separated in** $X \overset{\text{def}}{\Leftrightarrow} A, B \neq \emptyset \wedge A \cup B = X \wedge \bar{A} \cap B = A \cap \bar{B} = \emptyset)$, **so** $A \cap B = \emptyset$]

$\forall E \subseteq X(E$ disconnected $\overset{\text{def}}{\Leftrightarrow} \exists$ separated partition (A, B) of $E)$ | E connected $\Rightarrow \bar{E}$ connected ($\because$ take $(A \cap E, B \cap E)$ of E)
E connected $\not\Rightarrow E^\circ$ connected. Eg: $\bigcup 2$ open tangential disc. If $X = \mathbb{R}^1$, E° connected|E connected in $\mathbb{R}^k \Leftrightarrow E$ interval.

C1: Terminologies

Constrained NLP format: min/max $f(\vec{x})$ s.t. $\begin{cases} \forall i(g_i(\vec{x}) = 0) & \text{equality constraints} \\ \forall j(h_j(\vec{x}) \leq 0) & \text{inequality constraints} \end{cases}$

Unconstrained NLP format: min/max $f(\vec{x})$ s.t. $\vec{x} \in$ open convex D, f convex on D
Feasible region $S = \{\vec{x} \in \mathbb{R}^n : \forall i \forall j(g_i(\vec{x}) = 0 \wedge h_j(\vec{x}) \leq 0)\}$ closed set of feasible solutions (Generalisation)

Optimal $\vec{x}^* \in \mathbb{R}^k$(or ∞ if max or $-\infty$ if min) $\overset{\text{def}}{\Leftrightarrow} f(\vec{x})$ optimal objective value (or unbounded, so NLP unbounded).

C2: Weierstrass EVT, Graphical Method, Optimality Existence

Inner Product $\langle \vec{x}, \vec{y} \rangle = \vec{x}^T \vec{y} = \sum_{i=1}^{n} x_i y_i = \|\vec{x}\|\|\vec{y}\|\cos\theta_{xy}$ **(Cosine Rule) Properties**

1. $\langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$ [Additivity] | 2. $\langle k\vec{u}, \vec{v} \rangle = k\langle \vec{u}, \vec{v} \rangle$ [Homogeneity] | 3. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$ [Symmetry]
norm $\|\vec{u}\| = \langle \vec{u}, \vec{u} \rangle$ properties: 1. $\|\vec{u}\| \geq 0$, equality iff $\vec{u} = \vec{0}$ [Positive-Definiteness]
2. $|\|\vec{u}\| - \|\vec{v}\|| \leq \|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$ ($\because$ expand abs sign) [(Reverse) Triangle Inequality]
3. $|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|\|\vec{y}\| \Leftrightarrow \sum x_i y_i \leq \sum x_i^2 \sum y_i^2$, equality iff $\vec{x} // \vec{y}$ [Cauchy-Schwarz Inequality]

Vector Geometry [MA2104 Recap]

Scalar projection of $\vec{b}$ onto $\vec{a}$, $\text{comp}_{\vec{a}}\vec{b} = \|\vec{b}\|\cos\theta_{ab} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|}$|Vector projection of $\vec{b}$ onto $\vec{a}$, $\text{proj}_{\vec{a}}\vec{b} = \frac{\vec{a}}{\|\vec{a}\|}\text{comp}_{\vec{a}}\vec{b} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|^2}\vec{a}$

Cross-product: $\|\vec{a} \times \vec{b}\| = \|\vec{a}\|\|\vec{b}\||\sin\theta_{ab}|$ (Right-Hand Grip Rule) $\Rightarrow$ Perpendicular of $\vec{b}$ onto $\vec{a} = |\|\vec{b}\|\sin\theta_{ab}| = \frac{\|\vec{a} \times \vec{b}\|}{\|\vec{a}\|}$

Angle between planes=Angle between normals $= \cos^{-1}(\frac{\vec{n_1} \cdot \vec{n_2}}{\|\vec{n_1}\|\|\vec{n_2}\|}) \in [0, \pi]$|Arc length of curve $C = \vec{r}(t)$, $s = \int_a^b \|\vec{r'}(t)\|dt$

Extremisers [Maximisers' Definitions follow analogously]

Local Min: $\exists r > 0(\forall \vec{x} \in N_r(\vec{x}^*) \cap S)(f(\vec{x}) \geq f(\vec{x}^*))$|**Strict Local Min**: $\exists r > 0(\forall \vec{x} \in N_r(\vec{x}^*) \cap S \backslash \{\vec{x}^*\})(f(\vec{x}) > f(\vec{x}^*))$
Global Minimiser: $\forall \vec{x} \in S(f(\vec{x}) \geq f(\vec{x}^*))$|**Strict Global Min**: $\forall \vec{x} \in S \backslash \{\vec{x}^*\}(f(\vec{x}) > f(\vec{x}^*))$
(strict) Global maximisers $\Rightarrow$ (strict) Local maximisers. $\vec{x}^* \in E^\circ \Rightarrow \vec{x}^*$ stationary (not necessarily if $\in \partial E$)

Weierstrass Extreme Value Theorem (EVT), IVT, MVT

Weierstrass EVT: If f cont on nonempty closed bdd $S \subseteq \mathbb{R}^k$, $\exists$global minimiser $\wedge$ maximiser in S not-necessarily unique.
Boundedness Theorem: If f continuous on nonempty closed bounded $[a, b]$, f bounded in $[a, b]$.
$\because$ **IVT**: If real-valued f cont on closed bounded interval $[a, b] \neq \emptyset$, $\forall c \in (a, b)$, $\exists f^{-1}(c) \in (\min\{f(a), f(b)\}, \max\{f(a), f(b)\})$
$\therefore$ **Bolzano's Theorem**: If real-valued f continuous on closed bounded $[a, b] \neq \emptyset$ s.t. $f(a)f(b) < 0$, $\exists c \in (a, b)$
$\because$ **MVT**: If real-valued f cont on closed bounded interval $[a, b]$, diff on (a, b), $\exists c \in (a, b)$, $f'(c) = \frac{f(b)-f(a)}{b-a}$
$\therefore$ **Rolle's Theorem**: If real-valued f cont on $[a, b]$, diff on (a, b) s.t. $f(a) = f(b)$, $\exists c \in (a, b) f'(c) = 0$

C3: Convex/Non-Convex Set & Function

Convex Set: $\forall \vec{x}, \vec{y} \in S(\forall \lambda \in [0, 1])(\lambda\vec{x} + (1 - \lambda)\vec{y} \in S)$ **i.e, closure under affine combination**

Algebraic proofs oft-applies Traingle/Cauchy-Schwarz Inequalities for set/function. $\nexists$ "concave" analogue. Properties:
1. If $X = \mathbb{R}$ or $\mathbb{C}$, S convex $\Rightarrow S$ path-connected $\Rightarrow S$ connected. Converse fails. Eg: torus|2. ∂S convex curve/function.
3. $\emptyset, X$ convex set. $\forall S_1, S_2$ convex, countable intersection convex, but union convex iff chain (totally-ordered set).
Non-eg: Venn Diagram. ‖ 4. Strictly convex set: $\forall \vec{x}, \vec{y} \in S(\forall \lambda \in (0, 1))(\lambda\vec{x} + (1 - \lambda)\vec{y} \in {}^\circ)$.

Convex-downward Function: $\forall \vec{x}, \vec{y} \in D_f(\forall \lambda \in [0, 1])(f(\lambda\vec{x} + (1 - \lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1 - \lambda)f(\vec{y}))$ [$\exists$ **concave analogue**]

Strictly convex-downward function: $\forall \vec{x}, \vec{y} \in D_f(\forall \lambda \in (0, 1))(f(\lambda\vec{x} + (1 - \lambda)\vec{y}) < \lambda f(\vec{x}) + (1 - \lambda)f(\vec{y}))$. Properties:
1. $\forall f_1, \cdots, f_n$ convex, $\forall k_1, \cdots, k_n \geq 0$, $\sum_{i=1}^{n} k_i f_i$ convex [Linearity]. | 2. $\forall f_1$ convex, $\forall k < 0$, kf_1 concave.

3. $\forall f_1, f_2$ convex, max (min)$\{f_1, f_2\}$ convex (concave)|4. h convex (concave)$\wedge g \begin{cases} \text{non-dec} \\ \text{non-inc} \end{cases} \Rightarrow g \circ h \begin{cases} \text{convex (concave)} \\ \text{concave (convex)} \end{cases}$

C4: Gradient Vector, Tangent Plane

Convex Function f **on Convex Set** $S \subseteq \mathbb{R}^n$**:** $\forall k \in \mathbb{R}(S_k = \{\vec{x} \in D_f : f(\vec{x}) \leq k\})$ **convex, epigraph** E_f **convex**

$\forall \vec{x} = \sum \lambda_i \vec{x_i} \in D_f$, $\sum \lambda_i = 1$, $f(\vec{x}) \leq \sum \lambda_i f(\vec{x_i})$ ($\because$ Convex f's definition generalised by well-defined $f(\vec{x_i})$ in convex S)

Gradient Vector $\nabla f(\vec{x}) = \begin{pmatrix} f_{x_1}(\vec{x}) \\ \vdots \\ f_{x_n}(\vec{x}) \end{pmatrix} : \forall \vec{d} \in \mathbb{R}^n(\forall \lambda \in \mathbb{R})(f'(\vec{x} + \lambda\vec{d}) = (\nabla f(\vec{x} + \lambda\vec{d}))^T\vec{d})$

1. If $\lambda = 0$, $f'(\vec{x}) = \lim_{\lambda \to 0} \frac{f(\vec{x}+\lambda\vec{d})-f(\vec{x})}{\lambda} = (\nabla f(\vec{x}))^T\vec{d} =$ signed rate of change from $\vec{x}$ in direction $\vec{d}$
$\therefore f \downarrow$ most rapidly from $\vec{x}$ in direction $\vec{d} = -\nabla f(\vec{x})$ and $\uparrow$ most rapidly from $\vec{x}$ in direction $\vec{d} = \nabla f(\vec{x})$.
2. $\forall$ local min/maximiser $\vec{x}^*$ of f s.t. $g(\vec{x}) = c$ eq constraint, $\exists \lambda$ s.t.$\nabla f(\vec{x^*}) = \lambda \nabla g(\vec{x^*})$ [Lagrange Multiplier Method]
3. If f:convex $D_f \to \mathbb{R}^n$ convex, $\begin{cases} \vec{x}^* \text{ global minimiser} & \Leftrightarrow \forall \vec{x} \in D_f((\nabla f(\vec{x}^*)^T)(\vec{x} - \vec{x}^*) \geq 0) \ (\because f \uparrow \text{ in all dir from } \vec{x}^*) \\ \vec{x}^* \text{ global maximiser} & \Leftrightarrow \forall \vec{x} \in D_f((\nabla f(\vec{x}^*)^T)(\vec{x} - \vec{x}^*) \leq 0) \ (\because f \downarrow \text{ in all dir from } \vec{x}^*) \end{cases}$

Tangent-plane below f **Definition of Function Convexity**

f convex $\Leftrightarrow \forall \vec{x}, \vec{y} \in D_f(f(\vec{x}) + (\nabla f(\vec{x})^T)(\vec{y} - \vec{x}) \leq f(\vec{y}))$ [Strictly convex, concave, strictly concave apply analogously]

C5: Hessian Matrix, Definiteness Tests

Hessian $H_f(\vec{x}) = \begin{pmatrix} f_{x_1x_1} & f_{x_2x_1} & \cdots & f_{x_nx_1} \\ f_{x_1x_2} & f_{x_2x_2} & \cdots & f_{x_nx_2} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_1x_n} & f_{x_2x_n} & \cdots & f_{x_nx_n} \end{pmatrix} \forall \vec{x} \in {D_f}^\circ, (f_{x_i})_{x_j} = \frac{\partial^2 f}{\partial x_j x_i}, \Delta_k = \det(\begin{pmatrix} f_{x_1x_1} & f_{x_2x_1} & \cdots & f_{x_kx_1} \\ f_{x_1x_2} & f_{x_2x_2} & \cdots & f_{x_kx_2} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_1x_k} & f_{x_2x_k} & \cdots & f_{x_kx_k} \end{pmatrix}))$

Clairaut's Theorem: f has continuous 2nd-order partial derivatives $\Rightarrow H_f(\vec{x})$ symmetric. Else, non-eg is pathological.

3 Definiteness Tests

1. $\forall \vec{x} \in \mathbb{R}^n(\vec{x}^T A\vec{x} = \begin{cases} \geq 0 & \overset{\text{def}}{\Leftrightarrow} A \text{ p.s.d.} \\ > 0 & \overset{\text{def}}{\Leftrightarrow} A \text{ p.d.} \\ \leq 0 & \overset{\text{def}}{\Leftrightarrow} A \text{ n.s.d.} \\ < 0 & \overset{\text{def}}{\Leftrightarrow} A \text{ n.d.} \\ \exists \vec{x_1} \neq \vec{x_2}(\vec{x_1}^T A\vec{x_1} > 0 \wedge \vec{x_2}^T A\vec{x_2} < 0) & \overset{\text{def}}{\Leftrightarrow} A \text{ indefinite} \end{cases}$) [Definition]

2. $\forall$ eigenvalue λ of symmetric A, $\lambda = \begin{cases} \text{all } \geq 0 & \Leftrightarrow A \text{ p.s.d.} \\ \text{all } > 0 & \Leftrightarrow A \text{ p.d.} \\ \text{all } \leq 0 & \Leftrightarrow A \text{ n.s.d.} \\ \text{all } < 0 & \Leftrightarrow A \text{ n.d.} \\ \exists \lambda_1 < 0 \wedge \lambda_2 > 0 & \Leftrightarrow A \text{ indefinite} \end{cases}$ [Eigenvalue Test, $p_A(\lambda) = \det(A - \lambda I_n)$]

3. For symmetric A, $\begin{cases} \forall k \in [1, n]\Delta_k > 0 & \Leftrightarrow A \text{ p.d.} \\ \forall k \in [1, n](-1)^k\Delta_k > 0 & \Leftrightarrow A \text{ n.d.} \end{cases}$ [Principal Minor Test on Definiteness Only]

$\forall k \in [1, n]\Delta_k \geq 0 \not\Rightarrow A$ p.s.d. eg: n.s.d. $A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$. $\forall k \in [1, n](-1)^k\Delta_k \geq 0 \not\Rightarrow A$ n.s.d. eg: p.s.d. $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$

2nd Partial Derivative Test

f :open $D_f \to \mathbb{R}^n \begin{cases} \text{convex} & \Leftrightarrow \forall \vec{x}(H_f(\vec{x}) \text{ p.s.d.}) \\ \text{concave} & \Leftrightarrow \forall \vec{x}(H_f(\vec{x}) \text{ n.s.d.}) \\ \text{strictly convex} & \Leftarrow \forall \vec{x}(H_f(\vec{x}) \text{ p.d.}) \ [\text{Non-eg: } f(x) = x^4 \Rightarrow H_f(x) = 12x^2 \text{ p.s.d. only}] \\ \text{strictly convex} & \Leftarrow \forall \vec{x}(H_f(\vec{x}) \text{ n.d.}) \ [\text{Non-eg: } f(x) = -x^4 \Rightarrow H_f(x) = -12x^2 \text{ n.s.d. only}] \\ \text{neither} & \Leftarrow \exists \vec{x}(H_f(\vec{x}) \text{ indef}) \ [\text{Non-eg: } f(x) = x^3 \text{ neither by } (\nabla f)^T \Rightarrow H_f(x) = 6x \text{ p/n.s.d.}\forall x] \end{cases}$

Taylor's Theorem on $f : \mathbb{R}^n \to \mathbb{R}$ **with continuous 2nd partial derivatives**

$\forall \vec{x}, \vec{y} \in D_f([\vec{x}, \vec{y}] \subseteq {D_f}^\circ \Rightarrow \exists \lambda_w \in [0, 1](f(\vec{y}) = f(\vec{x}) + (\nabla f(\vec{x}))^T(\vec{y} - \vec{x}) + \frac{1}{2}(\vec{y} - \vec{x})^T H_f(\lambda_w\vec{x} + (1 - \lambda_w)\vec{y})(\vec{y} - \vec{x})))$

C6-C7: Function Coercivity, Critical Point Test of Semi-Definiteness

Coercive Function $f : \mathbb{R}^n \to \mathbb{R}$

Definition: $\lim_{\|\vec{x}\| \to \infty} f(\vec{x}) \to +\infty \overset{\text{def}}{\Leftrightarrow} \forall M > 0(\exists r > 0)(\forall \vec{x} \in N_r(\vec{0})^c)(f(\vec{x}) > M)$ i.e, all paths to $\pm\infty$ from $\vec{0}$ tends to $+\infty$
$\because \forall$ finite $p \in \mathbb{N}^+(\|\vec{x}\|_p = (\sum_{k=1}^{n} |x_k|^p)^{\frac{1}{p}} \Rightarrow \|\vec{x}\|_\infty = \lim_{p \to \infty} \|\vec{x}\|_p \cong \lim_{p \to \infty}(\max_{|x_i|}\{|x_i|^p\})^{\frac{1}{p}} = \underline{\max\{|x_i| : i \in [1, n]\}}$
$\therefore \|\vec{x}\| \to \infty \Leftrightarrow \|\vec{x}\|_\infty \to \infty$. Besides, $\forall \vec{x} \in \mathbb{R}^n(\|\vec{x}\|_\infty \leq \|\vec{x}\| \leq n\|\vec{x}\|_\infty)$.
f **coercive** $\Rightarrow \exists$ **finite global minimiser** ($\because$ **Weierstrass EVT on closed ball** $\overline{N_r(\vec{0})}$). Converse non-eg: $f(x) = \frac{x^2}{1+x^2}$
$\therefore$ If coercive f has local minimisers, global $\vec{x}^*$ must be within.

Interior Critical Point Tests

$\vec{x}^* \begin{cases} \text{local minimiser} & \Rightarrow \nabla f(\vec{x}^*) = \vec{0} \wedge H_f(\vec{x}^*) \text{ p.s.d. } [\text{Non-eg: } x = 0, f(x) = -x^4] \\ \text{local maximiser} & \Rightarrow \nabla f(\vec{x}^*) = \vec{0} \wedge H_f(\vec{x}^*) \text{ n.s.d. } [\text{Non-eg: } x = 0, f(x) = x^4] \\ \text{strict local minimiser} & \Leftarrow \nabla f(\vec{x}^*) = \vec{0} \wedge H_f(\vec{x}^*) \text{ p.d. } [\text{Non-eg: } x = 0, f(x) = x^4 \text{ with } H_f(0) = 0 \text{ p.s.d. only}] \\ \text{strict local maximiser} & \Leftarrow \nabla f(\vec{x}^*) = \vec{0} \wedge H_f(\vec{x}^*) \text{ n.d. } [\text{Non-eg: } x = 0, f(x) = -x^4 \text{ with } H_f(0) = 0 \text{ n.s.d. only}] \\ \text{saddle point} & \Leftarrow \nabla f(\vec{x}^*) = \vec{0} \wedge H_f(\vec{x}^*) \text{ indef } [\text{Non-eg: } f(x, y) = x^3 - y^3, H_f(0, 0) = (0) \text{ singular p/n.s.d}] \end{cases}$
f coercive $\wedge \partial D_f = \emptyset$ ($\because$ unconstrained) $\Rightarrow$ any global minimiser is interior critical point, saving sufficient ∂D_f analysis.

Convex NLP format: min/max $f(\vec{x})$ s.t. $\vec{x} \in$ convex $D \neq \emptyset$, f convex on D [D non-open if constrained]

$\vec{x}^*$ local minimiser (maximiser) $\wedge$f convex (concave) $\Rightarrow \vec{x}^*$ global minimiser (maximiser) [$\vec{x}^*$ strict$\Leftrightarrow$ f strict]

Unconstrained Convex Quadratic Programming format

min/max $q(\vec{x}) = \frac{1}{2}\vec{x}^T Q\vec{x} + \vec{c}^T\vec{x}$, Q **symmetric s.t.** $\vec{x} \in$ **open convex** $D \neq \emptyset$, q **convex on** D. Properties:

1. $q(\vec{x})$ convex on D $\Leftrightarrow$ Q p.s.d. ($\because \nabla q(\vec{x}) = Q\vec{x} + \vec{c} \Rightarrow H_q(\vec{x}) = Q$, 2nd Partial Derivative Test)
2. $q(\vec{x})$ convex (concave) on D $\Rightarrow$ ($\vec{x}^*$ local (hence global) minimiser (maximiser) $\Leftrightarrow Q\vec{x}^* = -c$)
3. $\inf_{\vec{x}\in\mathbb{R}^n}\{\frac{1}{2}\vec{x}^TQ\vec{x} - \vec{c}^T\vec{x}\} = \begin{cases} -\frac{1}{2}\vec{w}^TQ\vec{w} & \forall\vec{c} \in Col(Q) \overset{def}{\Leftrightarrow} \exists\vec{w} \in \mathbb{R}^n(\vec{c} = Q\vec{w})[\text{Non-unique } \vec{w} \Leftrightarrow \text{Nullity}(Q) > 0] \\ -\infty & \forall\vec{c} \notin Col(Q)\ (\therefore Q \text{ deficient-rank}) \end{cases}$

Above infinum takes Q symmetric p.s.d so LHS convex, $-\frac{1}{2}\vec{w}^TQ\vec{w}$ unique.

C8: 3 Numerical Methods to 1-variable Unconstrained NLP

Bisection (aka Binary Search) Method ($\because$ Bolzano's Theorem)

Necessary conditions: f continuous, 1-differentiable

1. Pick interval $[a_1, b_1]$ s.t. $f'(a_1)f'(b_1) \leq 0$, and width tolerance $\epsilon > 0$.
2. if $f'(a_1) = 0$: `return` a_1. if $f'(b_1) = 0$: `return` b_1.
3. while $b_k - a_k > 2\epsilon$:
4. med=$a_k + \frac{b_k - a_k}{2}$
5. if $f'(\text{med}) = 0$: return med.
6. if $f'(\text{med})f'(b_k) < 0$: $a_{k+1} = \text{med}; b_{k+1} = b_k$
7. else $b_{k+1} = \text{med}; a_{k+1} = a_k$
8. `return` $a_f + \frac{b_f - a_f}{2}$

Convergence: $b_k - a_k = \frac{b_1 - a_1}{2^{k-1}} \Rightarrow f \leq \lceil\log_2(\frac{b_1-a_1}{\epsilon})\rceil \Leftrightarrow$ max steps $= \lceil\log_2(\frac{b_1-a_1}{\epsilon})\rceil$ ($\because$ 1-indexing)

Newton's Method ($\because$ Truncated Taylor's Expansion)

Necessary conditions: f continuous, 2-differentiable

$\because f(x) \approx q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2 = \frac{1}{2}f''(x_k)[x - x_k + \frac{f'(x_k)}{f''(x_k)}]^2 + [f(x_k) - \frac{(f'(x_k))^2}{2f''(x_k)}]$

$\therefore$ Newton's univariate recurrence: $x^* \approx x_{k+1} \approx x_k - \frac{f'(x_k)}{f''(x_k)}$

1. Pick x_0, derivative tolerance $\epsilon > 0$.
2. while $|f'(x_k)| > \epsilon : x_{k+1} = x_k - (f''(x_k))^{-1}f'(x_k)$
3. `return` x_f

Convergence for close x_0 to x^*: quadratic.

Golden Section Method ($\because$ Unimodality)

Necessary conditions: f continuous, unimodal

1. Pick interval $[a_1, b_1]$ s.t. $\exists!$ local minimiser $x^* \in [a_1, b_1]$, and width tolerance $\epsilon > 0$. Cache golden ratio $\phi = \frac{\sqrt{5}-1}{2}$
2. $\ell_1 = \phi a_1 + (1 - \phi)b_1; h_1 = (1 - \phi)a_1 + \phi b_1$
3. while $b_k - a_k > \epsilon$:
4. if $f(\ell_k) > f(h_k) : a_{k+1} = \ell_k; b_{k+1} = b_k; h_{k+1} = \ell_k + \phi(b_k - \ell_k); \ell_{k+1} = h_k$ ($\therefore 4^{th}$ assignment: cost-saving narrowing)
5. else: $b_{k+1} = h_k; a_{k+1} = a_k; \ell_{k+1} = h_k - \phi(h_k - a_k); h_{k+1} = \ell_k$
6. `return` $\frac{a_f + b_f}{2}$

Convergence: $b_k - a_k = \phi^{k-1}(b_1 - a_1) \Rightarrow f \leq \lceil\log_{\phi^{-1}}(\frac{b_1-a_1}{\epsilon})\rceil \Leftrightarrow$ max steps $= \lceil\log_{\phi^{-1}}(\frac{b_1-a_1}{\epsilon})\rceil$ ($\because$ 1-indexing)

C9-C11: 3 Numerical Methods to ¿1-variable Unconstrained NLP

Multivariable Newton's Method

4 Necessary conditions: f continuous, 2-differentiable, $\forall k \in \mathbb{N}_0(H_f(\vec{x}^{(k)})$ invertible$), \vec{x}^{(0)}$ close to unknown $\vec{x}^*$
$H_f(\vec{x}^{(k)})$ Lipschitz-continuous: $(\exists r, L > 0)(\forall\vec{x}, \vec{y} \in N_r(\vec{x}^*))(\|H_f(\vec{x}) - H_f(\vec{y})\| \leq L\|\vec{x} - \vec{y}\|)$

1. Pick $\vec{x}^{(0)}$, and derivative tolerance $\epsilon > 0$.
2. while $\|\nabla f(\vec{x}^{(k)})\| \geq \epsilon$: $\vec{x}^{(k+1)} = \vec{x}^{(k)} - H_f(\vec{x}^{(k)})^{-1}\nabla f(\vec{x}^{(k)})$
3. `return` $\vec{x}^{(f)}$

Convergence for close $\vec{x}^{(0)}$ to $\vec{x}^*$: quadratic. Optionally improve via Armijo line search:

1. $\epsilon_{arm} \in (0, 0.5); r \in (0, 1); t = 1$ usually
2. while $f(\vec{x}^{(k)} + t\vec{d}^{(k)}) > f(\vec{x}^{(k)}) + \epsilon_{arm}t\nabla f(\vec{x}^{(k)})^T\vec{d}^{(k)}$: $t = t * r$ [Generalisable to other methods' $\vec{d}^{(k)}$]

Steepest Descent Method

Necessary conditions: f continuous, 1-differentiable

1. Pick $\vec{x}^{(0)}$, and derivative tolerance $\epsilon > 0$
2. while $\|\nabla f(\vec{x}^{(k)})\| \geq \epsilon$: $\vec{x}^{(k+1)} = \vec{x}^{(k)} - t_k\nabla f(\vec{x}^{(k)}), t_k = \text{argmin}_{t\geq 0}g(\vec{x}^{(k)} - t_k\nabla f(\vec{x}^{(k)}))$ default line search.
3. `return` $\vec{x}^{(f)}$

Convergence for quadratic $q(\vec{x}) := \frac{1}{2}\vec{x}^TQ\vec{x} - \vec{c}^T\vec{x}$, $\rho(Q) := (\frac{\kappa(Q)-1}{\kappa(Q)+1})^2 = 1 - \frac{4}{\kappa(Q)+1} + \frac{4}{(\kappa(Q)+1)^2} \approx 1 - \frac{4}{\kappa(Q)}$

Matrix condition $\kappa(Q) := \frac{\lambda_{max}(Q)}{\lambda_{min}(Q)} = (\frac{\text{major axis}}{\text{minor axis}})^2 \Rightarrow$ step $= \lfloor\frac{\ln\epsilon}{\ln\rho(Q)}\rfloor + 1$. Step length $t_k = \frac{\|\nabla f(\vec{x}^{(k)})\|^2}{\langle\nabla f(\vec{x}^{(k)}), Q\nabla f(\vec{x}^{(k)})\rangle}$

Conjugate Descent Method only for convex quadratic programming $\min f(\vec{x}) := \frac{1}{2}\vec{x}^TQ\vec{x} - \vec{c}^T\vec{x}$

Necessary conditions: Q symmetric p.d ($\therefore$ invertible).

1. Pick $\vec{x}^{(0)}$. Exact method lacks ϵ. Compute initial residual $\vec{r}^{(0)} := -\nabla f(\vec{x}^{(0)}) := Q\vec{x}^{(0)} - \vec{c}; \vec{p}^{(0)} = -\vec{r}^{(0)}$.
2. while $\vec{r}^{(0)} \neq \vec{0}$: $\vec{x}^{(k+1)} = \vec{x}^{(k)} + t_k\vec{p}^{(k)} = \vec{x}^{(k)} - \frac{\langle\vec{r}^{(k)}, \vec{p}^{(k)}\rangle}{\langle\vec{p}^{(k)}, Q\vec{p}^{(k)}\rangle}\vec{p}^{(k)}$
3. $\vec{r}^{(k+1)} = \nabla f(\vec{x}^{(k+1)}) = Q\vec{x}^{(k+1)} - \vec{c} = \vec{r}^{(k)} + t_kQ\vec{p}^{(k)} = \vec{r}^{(k)} - \frac{\langle\vec{r}^{(k)}, \vec{p}^{(k)}\rangle}{\langle\vec{p}^{(k)}, Q\vec{p}^{(k)}\rangle}Q\vec{p}^{(k)}$
4. $\vec{p}^{(k+1)} = -\vec{r}^{(k+1)} + \beta_{k+1}\vec{p}^{(k)} = -\vec{r}^{(k+1)} + \frac{\langle\vec{r}^{(k+1)}, Q\vec{p}^{(k)}\rangle}{\langle\vec{p}^{(k)}, Q\vec{p}^{(k)}\rangle}\vec{p}^{(k)} = -\vec{r}^{(k+1)} + \frac{\langle\vec{r}^{(k+1)}, \vec{r}^{(k+1)}\rangle}{\langle\vec{r}^{(k)}, \vec{r}^{(k)}\rangle}\vec{p}^{(k)}$

Properties: $\{\vec{p}^{(0)} \cdots \vec{p}^{(n)}\}$ conjugate wrt Q $\overset{def}{\Leftrightarrow} (\forall i \neq j)(\langle\vec{p}^{(i)}, Q\vec{p}^{(j)}\rangle = 0)$. Conjugacy non-transitive, symmetric iff Q is.
$\vec{r}^{(k+1)} = \vec{r}^{(0)} + \sum_{i=0}^{k} t_i\vec{p}^{(i)}$ inductively $\Rightarrow (\forall i < k)(\langle\vec{r}^{(k)}, \vec{p}^{(i)}\rangle = 0)$ given $\vec{p}^{(0)} = -\vec{r}^{(0)}$ base case.

Convergence: $\vec{x}^{(k)} = \vec{x}^{(0)} + \sum_{i=0}^{k-1} t_i\vec{p}^{(i)} \in \vec{x}^{(0)} + \text{Span}\{\vec{p}^{(0)} \cdots \vec{p}^{(k-1)}\} \wedge \vec{p}^{(i)}$'s L.I. $\Rightarrow \vec{x}^{(n)}$ min f in $\mathbb{R}^n \Rightarrow$ exact steps $\leq$ n

C12-C13: Regular & KKT Points

Regular Points $\vec{x}^*$:= feasible points fulfilling Linear Independence Constraint Qualification (LICQ)

LICQ: Active constraint gradient set (Constraint Jacobian) $A(\vec{x}^*) = \{\nabla g_i(\vec{x}^*)\} \cup \{\nabla h_j(\vec{x}^*) : j \in J(\vec{x}^*)$ active indices$\}$ LI

Corollary: $\nexists g_i \Rightarrow$ all interior points are vacuously-regular ($J(\vec{x}^*) = \emptyset$ if all inequalities inactive/slack)

Karush-Kuhn-Tucker (KKT) points $\vec{x}^*$ Defintion

1. Stationarity: ($\exists$ Lagrangian multiplier λ_i^*, μ_j^*'s)(Lagrangian $L_{\vec{x}}(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) := f(\vec{x}^*) + \sum\lambda_i^*g_i(\vec{x}^*) + \sum\mu_j^*h_j(\vec{x}^*) = \vec{0}$)
2. Primal feasibility: $(\forall i)(g_i(\vec{x}^*) = \vec{0}) \wedge (\forall j)(h_j(\vec{x}^*) = \vec{0})$| 3. Dual feasibility: $(\forall j)(\mu_j^* \geq 0)$ from stationarity
4. Complementary slackness: $(\forall j)(\mu_j^*h_j(\vec{x}^*) = \vec{0} \Leftrightarrow h_j(\vec{x}^*)$ active $\vee\mu_j^* = 0\forall h_j(\vec{x}^*)$ inactive)

Strict complementarity (not needed): $(\forall h_j(\vec{x}^*)$ active$)(\mu_j^* > 0)$

KKT 1ONC: $\vec{x}^*$ regular $\wedge$ local min $\Rightarrow \vec{x}^*$ KKT

Analogous unconstrained 1st-Order Partial Derivative Test ($\therefore \vec{x}^*$ interior): $\vec{x}^*$ local minimiser $\Rightarrow \vec{x}^*$ critical point

Critical point $\vec{x}^*$: $\nabla f(\vec{x}^*) = \vec{0}\vee$ undefined | $\nexists$ local constrained general tests $\Rightarrow$ KKT 1ONC hard to apply.

KKT 2ONC: $\vec{x}^*$ regular $\wedge$ local min $\wedge f, g_i, h_j$ 2-diff $\Rightarrow \vec{x}^*$ KKT $\wedge(\forall\vec{y} \in C(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*))(H_L(\vec{x}^*)$ p.s.d.)

Critical cone $C(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) := \{\vec{y} \in \mathbb{R}^n : \begin{cases} \langle\nabla g_i(\vec{x}^*), \vec{y}\rangle = 0 & \\ \langle\nabla h_j(\vec{x}^*), \vec{y}\rangle = 0 & \forall \text{ active } h_j(\vec{x}^*), \mu_j^* > 0 \\ \langle\nabla h_j(\vec{x}^*), \vec{y}\rangle \leq 0 & \forall \text{ active } h_j(\vec{x}^*), \mu_j^* = 0 \end{cases}$ independent of inactive $h_j(\vec{x}^*)\}$

$C(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) \subseteq$ Linearised feasible direction set $F(\vec{x}^*) := \{\vec{y} \in \mathbb{R}^n : \langle\nabla g_i(\vec{x}^*), \vec{y}\rangle = 0 \wedge \langle\nabla h_j(\vec{x}^*), \vec{y}\rangle \leq 0\forall$ active $h_j(\vec{x}^*)\}$

Analogous unconstrained 2nd-Order Partial Derivative Test ($\therefore \vec{x}^*$ interior): $\vec{x}^*$ local minimiser $\wedge$f 2-diff $\Rightarrow H_f(\vec{x}^*)$ p.s.d.

Corollary: X regular throughout, or not for finite points $\Rightarrow$ global ($\because$ local) minimiser $\vec{x}^*$ KKT or in finite points.

KKT 2OSC: $\vec{x}^*$ KKT $(\forall\vec{y} \in C(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*))(H_L(\vec{x}^*)$ p.d.) $\Rightarrow \vec{x}^*$ strict local minimiser, $H_L(\vec{x}^*) = \nabla_{\vec{x}\vec{x}}L(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*)$

Analog unconstrained 2nd-Order Partial Derivative Test ($\therefore \vec{x}^*$ interior): $\nabla f(\vec{x}^*) = \vec{0} \wedge H_f(\vec{x}^*)$ pd $\Rightarrow \vec{x}^*$ strict local min

Farka's Lemma: $(\forall A \in M_{m\times n}(\mathbb{R}))(\forall\vec{b} \in \mathbb{R}^m)((\exists\vec{x} \in \mathbb{R}^n)(A\vec{x} = \vec{b} \wedge \vec{x} \geq \vec{0}$ **item-wise** $) \oplus (\exists\vec{y} \in \mathbb{R}^m)(\vec{b}^T\vec{y} < \vec{0} \leq A^T\vec{y}))$

C14-C15: Constrained Convex NLP

Convex Implications and Slater's Condition

Constrained convex NLP format: $\min_{\vec{x}\in\text{convex } X}$ convex $f(\vec{x})$ s.t. $\begin{cases} g_i(\vec{x}) := \vec{a_i}^T\vec{x} - b_i = 0 \text{ affine/linear} \\ h_j(\vec{x}) \leq 0 \text{ convex} \\ \text{Feasible set } X \text{ convex} \end{cases}$

Slater's Condition: $(\exists$ strictly-interior $\hat{\vec{x}} \in X)(g_i(\hat{\vec{x}}) = 0 \wedge h_j(\hat{\vec{x}}) < 0)$

Corollaries: 1. $\vec{x}^*$ KKT in convex program $\Rightarrow \vec{x}^*$ global minimiser
2. $\vec{x}^*$ global-minimiser in convex program $\wedge$ Slater's $\Rightarrow \vec{x}^*$ KKT. $\therefore$ Iff holds
Linear convex ECP: $J(\vec{x}^*) = \emptyset \Rightarrow$ vacuous LICQ $\wedge$ Slater's $\Rightarrow (\vec{x}^*$ local min $\Leftrightarrow \vec{x}^*$ KKT $\Leftrightarrow \vec{x}^*$ global min))
$\because\Rightarrow$: KKT 1ONC, convexity, $\Leftarrow$: Global, Slater's
3. Linear general ECP ($X \neq \emptyset$), $\vec{x}$ KKT $\Leftrightarrow \vec{x}^*$ global min. $\therefore$ Suffice to check stationarity.

Orthogonality

$\vec{u} \perp \vec{v} \overset{def}{\Leftrightarrow}$ Standard $\langle\vec{u}, \vec{v}\rangle = \vec{u}^T\vec{v} = 0$| subspaces $U \perp V \overset{def}{\Leftrightarrow} (\forall\vec{u} \in U)(\forall\vec{v} \in V)(\vec{u} \perp \vec{v})$. Eg: $Im(A^T) \perp Ker(A)$ provably.

Constrained (convex non-ECP) quadratic programming: min $q(\vec{x}) := \frac{1}{2}\vec{x}^TQ\vec{x} + \vec{c}^T\vec{x}$ **s.t.** $A\vec{x} - \vec{b} \leq \vec{0} \wedge \vec{x} \geq \vec{0}$

Support Vector Machine (SVM): max $\frac{2}{\|\vec{w}\|} \equiv \min f(\vec{w}) = \frac{1}{2}\|\vec{w}\|^2$ s.t. $h_j(\vec{w}) := -y_j(\vec{x_j}^T\vec{w} + b) + 1 \leq 0, y_j \in \{\pm 1\}$

$\therefore Q = I_n$ symmetric p.d. $\wedge\vec{c} = \vec{0}$. $\vec{x}^*$ KKT in convex QP SVM $\Rightarrow \vec{x}^*$ global-minimiser, hence find global minimum.

Hinge-loss SVM: $\min f(\vec{w}) = \frac{1}{2}\|\vec{w}\|^2 + \lambda\sum\zeta_i$ s.t. $h_j(\vec{w}) := -y_j(\vec{x_j}^T\vec{w} + b) - \zeta_j + 1 \leq 0 \wedge h_k(\vec{w}) := -\zeta_k \leq 0, y_j \in \{\pm 1\}$

Constrained (convex) LP: min $f(\vec{x}) := \vec{c}^T\vec{x}$ **s.t.** $g_i(\vec{x}) := b_i - a_i^Tx_i = 0 \wedge h_j(\vec{x}) := -e_j^Tx_j \leq 0$ ($\therefore \vec{x} \geq 0$)

C16-C17: Lagrangian Dual Problem

Lagrangian Dual Formulation

$\because$ Primal minimisation format: $\min_{\vec{x}\in X}\{\max_{\vec{\lambda}\in\mathbb{R}^m, \vec{\mu}\geq\vec{0}_n}\{L(\vec{x}, \vec{\lambda}, \vec{\mu}) := f(\vec{x}) + \sum\lambda_ig_i(\vec{x}) + \sum\mu_jh_j(\vec{x})\}\}$

$\therefore$ Swap: Lagrangian dual function $\theta(\vec{\lambda}, \vec{\mu}) := \inf_{\vec{x}\in X} L(\vec{x}, \vec{\lambda}, \vec{\mu}) \Rightarrow$ Lagrangian dual problem $D := \max_{\vec{\lambda}\in\mathbb{R}^m, \vec{\mu}\geq\vec{0}_n}\theta(\vec{\lambda}, \vec{\mu})$

Corollaries: 1. θ concave independent of f, g_i, h_j's convexity.
2. If single IECP, graphically, primal $:= \min_{x\leq 0} y \geq$ dual $:= \min$ y-intercept of downward-sloping tangent

Duality Theorems

Weak Duality Theorem: ($\forall$ primal feasible $\vec{x}$)($\forall$ dual feasible $(\vec{\lambda}, \vec{\mu})$)($f(\vec{x}) \geq \theta(\vec{\lambda}, \vec{\mu}) \Leftrightarrow$ duality gap ≥ 0)

Strong Duality Theorem: X convex set $\wedge f, h_j$ convex $\wedge g_i$ affine/linear $\wedge$ Slater's condition $\Rightarrow$ P $-$ D :=duality gap $= 0$

$\because$ If $f(\vec{x}^*) = \theta(\vec{\lambda}^*, \vec{\mu}^*)$, then $L(\vec{x}, \vec{\lambda}^*, \vec{\mu}^*) = \inf_{\vec{x}\in X} L(\vec{x}, \vec{\lambda}^*, \vec{\mu}^*) := \theta(\vec{\lambda}^*, \vec{\mu}^*) = f(\vec{x}^*) = L(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) \Rightarrow (\vec{\mu}^*)^Th(\vec{x}^*) = 0$.
$\therefore$ P = D $\Rightarrow$ complementary slackness.

Convex proximal-mapping format P: $\min_{B\vec{x}=\vec{u}} f(\vec{x}) := \frac{1}{2}\|\vec{x} - \vec{c}\|^2 + \rho\|\vec{u}\|_\infty$

$\therefore D := \max_{\vec{\lambda}\in\mathbb{R}^m, \sum|\lambda|\leq\rho}\theta(\vec{\lambda}) := -\frac{1}{2}\|B^T\vec{\lambda}\|^2 + \langle B\vec{c}, \vec{\lambda}\rangle$

C18: Saddle Point of Lagrangian function $L(\vec{x}, \vec{\lambda}, \vec{\mu})$

Definition: $(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*)$ saddle-point $\overset{def}{\Leftrightarrow} (\forall\vec{x} \in X)(\forall(\vec{\lambda}, \vec{\mu}) \in \mathbb{R}^m \times \mathbb{R}^n_+)(L(\vec{x}^*, \vec{\lambda}, \vec{\mu}) \leq L(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) \leq L(\vec{x}, \vec{\lambda}^*, \vec{\mu}^*))$

Corollaries: 1. $(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*)$ saddle-point $\Rightarrow \vec{x}^*$ primal global minimiser $\wedge(\vec{\lambda}^*, \vec{\mu}^*)$ dual global maximiser.
2. $(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*)$ saddle-point $\wedge\vec{x}^* \in X^\circ \Rightarrow (\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*)$ KKT ($\because \nabla_{\vec{x}}L(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) = \vec{0}$)
3. $\vec{x}^*$ KKT-point of convex program with Lagrangian $\vec{\lambda}^*, \vec{\mu}^* \Rightarrow (\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*)$ saddle-point.
4. $f(\vec{x}^*) > \theta(\vec{\lambda}^*, \vec{\mu}^*) \Rightarrow \nexists$ saddle.

C19-C22: 5 Numerical Methods to $\geq$ 1-variable Constrained NLP

Subdifferential Terminologies & Properties

Subdifferential $\partial f(\vec{x}^*) := \{\text{subgradient } \zeta : (\forall \vec{x} \in X)(\begin{cases} f(\vec{x}) \leq f(\vec{x}^*) + \zeta^T(\vec{x} - \vec{x}^*) & \forall \text{concave } f(\text{eg: } \theta) \\ f(\vec{x}) \geq f(\vec{x}^*) + \zeta^T(\vec{x} - \vec{x}^*) & \forall \text{convex } f \end{cases})\}$ convex set

$\because$ Constraint vector $\beta(\vec{x}) := (g(\vec{x}), h(\vec{x}))$, Lagrangian multiplier vector $\vec{w} := (\vec{\lambda}, \vec{\mu})$

$\therefore$ ($\forall$ Lagrangian minimiser $\vec{x}^* \in X(\vec{w}))(\theta(\vec{w}) = L(\vec{x}^*, \vec{w}^T\beta(\vec{x}^*)))$. Corollaries:

1. If $X(\vec{w}) = \{\vec{x}^*\}$ singleton, then $\nabla\theta(\vec{w}) = \beta(\vec{x}^*)$ well-defined at $\vec{w}$ | 2. $\partial\theta(\vec{w}) = \text{conv}\{\beta(\vec{x}^*) : \vec{x}^* \in X(\vec{w})\}$

3. $(\exists \vec{x}^* \in X(\vec{w}))(\vec{d}^T\beta(\vec{x}^*) \leq$ r.o.c. of θ in direction $\vec{d}$ = directional derivative $\theta'(\vec{w};\vec{d}) := \lim_{\gamma\to 0^+} \frac{\theta(\vec{w}+\gamma\vec{d})-\theta(\vec{w})}{\gamma})$

(3)'s continuous analogy: $\vec{d}^T\nabla f(\vec{x}) =$ r.o.c. of f in direction $\vec{d}$| $\therefore \theta'(\vec{w};\vec{d}) = \inf_{\zeta\in\partial\theta(\vec{w})}\{\vec{d}^T\zeta\}$

$\because \vec{d}$ ascent direction of θ at $\vec{w} \overset{\text{def}}{\Leftrightarrow} (\exists r > 0)(\forall t \in (0, r))(\theta(\vec{w} + t\vec{d}) > \theta(\vec{w})) \overset{\text{def}}{\Leftrightarrow} (\exists r > 0)(\forall \zeta \in \partial\theta(\vec{w}))(\vec{d}^T\zeta \geq r)$

$\therefore$ Normalised steepest ascent direction of θ at $\vec{w}$, $\vec{d} := \frac{\hat{\zeta}}{\|\hat{\zeta}\|}, \hat{\zeta} := \min\{\zeta \in \partial\theta(\vec{w}) = \text{conv}\{\beta(\vec{x}^*) : \vec{x}^* \in X(\vec{w})\}\}$

$\therefore$ Ascent directions' cone $C(\vec{d}) := \{\vec{d} : \inf_{\zeta\in\partial\theta(\vec{w})}\{\vec{d}^T\zeta\} \geq 0\}$. Graphically, $\forall\zeta \in \partial\theta(\vec{w})$, angle between ζ and $\theta \leq \frac{\pi}{2}$.

Steepest Ascent Direction Method for $\max_{\vec{w}\in\mathbb{R}^m\times\mathbb{R}^p_+} \theta(\vec{w}) := \inf_{\vec{x}\in X}\{f(\vec{x}) + \vec{w}^T\beta(\vec{x})\}$ dual maximisation

1. Pick $\vec{w}^{(0)}$, and min-norm tolerance $\epsilon > 0$. Compute min-norm $\hat{\zeta}$ as below.
2. while $\frac{1}{2}\|\hat{\zeta}\|^2 > \epsilon$: subdifferential $\partial\theta(\vec{w}^{(k)}) = \text{conv}\{\beta(\vec{x}^*) : \vec{x}^* \in X(\vec{w}^{(k)})\}$
3. normalised steepest ascent direction $\vec{d}^{(k)} = \frac{\hat{\zeta}}{\|\hat{\zeta}\|}, \hat{\zeta} = \text{argmin}_{\zeta\in\partial\theta(\vec{w}^{(k)})}\{\frac{1}{2}\|\zeta\|^2\}$
4. $\vec{w}^{(k+1)} = \vec{w}^{(k)} + t_k\vec{d}^{(k)}, t_k = \text{argmin}_{t\geq 0}\{\vec{w}^{(k)} + t\vec{d}^{(k)}\}$
5. `return` $\vec{w}^{(f)}$. ($\therefore$ if steepest descent, only edit Step 3 to $-\vec{d}^{(k)}$)

Frank-Wolfe Algorithm for convex IECP $\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x})$ s.t. $A\vec{x} \leq \vec{b}$

1. Pick $\vec{x}^{(0)}$, and width tolerance $\epsilon > 0$. Set $\ell_0 = -\infty, h_0 = f(\vec{x}^{(0)})$.
2. while $h_k - \ell_k > \epsilon$: solve $\hat{\vec{x}}^{(k)} = \text{argmin } z(\vec{x}) = f(\vec{x}^{(k)}) + \nabla f(\vec{x}^{(k)})^T(\vec{x} - \vec{x}^{(k)})$ s.t. $A\vec{x} \leq \vec{b}$ constrained linear subproblem.
3. $\ell_{k+1} = \max(\ell_k, z(\hat{\vec{x}}^{(k)})) = \max\{\ell_0 \cdots \ell_k, z(\hat{\vec{x}}^{(k)})\}$ nondecreasing.
4. Optional line search: feasible direction $\vec{d}^{(k)} := \hat{\vec{x}}^{(k)} - \vec{x}^{(k)} \Rightarrow t_k = \text{argmin}_{t\in[0,1]}\{f(\vec{x}^{(k)} + t\vec{d}^{(k)})\}$
5. $\vec{x}^{(k+1)} = \vec{x}^{(k)} + t_k\vec{d}^{(k)}; h_{k+1} = f(\vec{x}^{(k+1)})$ ($\therefore h_i$ may fluctuate)
6. `return` $\vec{x}^{(f)}$. Corollaries: 1. f convex $\Rightarrow$ Taylor's approximate $z(\vec{x}) \leq f(\vec{x})$|2. $\vec{x}^{(k)}$ minimises f $\Leftrightarrow \vec{x}^{(k)}$ minimises $z_k(\vec{x})$

Quadratic Penalty Method for (not-necessarily linear/convex) ECP $\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x})$ s.t. $g(\vec{x}) = \vec{0}$

1. Pick $\vec{x}^{(0)}$, and constraints tolerance $\epsilon > 0$. Conventionally set initial penalty $\mu_0 = 1$.
2. Crucially formulate $\min_{\vec{x}\in\mathbb{R}^n} Q(\vec{x};\mu_k) := f(\vec{x}) + \frac{1}{2\mu_k}\sum g_i^2(\vec{x})$ unconstrained subproblem.
3. while $\|g(\vec{x}^{(k)})\| \geq \epsilon$: solve $\vec{x}^{(k+1)} = \text{argmin}Q(\vec{x};\mu_k)$ unconstrained subproblem, oft-via same ϵ and Newton's Method.
4. $\mu_{k+1} = \rho\mu_k$ for some $\rho \in (0, 1)$ i.e, penalise positive constraint values more.
5. `return` $\vec{x}^{(f)}$.

Corollaries: 1. $(\forall k \in \mathbb{N}_0)(\vec{x}^{(k+1)}$ exactly min $Q(\vec{x};\mu_k)$ at $\epsilon_k = 0) \wedge \mu_k \to 0^+ \Rightarrow \{\vec{x}^{(k)}\} \to \vec{x}^*$ exactly min f.

2. If $(\forall k \in \mathbb{N}_0)(\vec{x}^{(k+1)}$ approx min $Q(\vec{x};\mu_k)$ s.t. $\|\nabla_{\vec{x}}Q(\vec{x}^{(k+1)};\mu_k)\| \leq \epsilon_k \to 0^+) \wedge \mu_k \to 0^+$, then $\{\vec{x}^{(k)}\} \to \vec{x}^*$ regular feasible $\Leftrightarrow \vec{x}^*$ KKT with $\lambda_i^* = \lim_{k\to\infty}\frac{g_i(\vec{x}^{(k+1)})}{\mu_k}$. Note $\vec{x}^*$ may not be regular/feasible.

Augmented Lagrangian Method for (not-necessarily linear/convex) ECP $\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x})$ s.t. $g(\vec{x}) = \vec{0}$

1. Pick $\vec{x}^{(0)}$, and constraints tolerance $\epsilon > 0$. Conventionally set initial penalty $\mu_0 = 1$.
2. Crucially formulate $\min_{\vec{x}\in\mathbb{R}^n} L_A(\vec{x};\vec{\lambda}^{(k)},\mu_k) := f(\vec{x}) + \sum\lambda_i^{(k)}g_i(\vec{x}) + \frac{1}{2\mu_k}\sum g_i^2(\vec{x})$ unconstrained subproblem.
3. while $\|g(\vec{x}^{(k)})\| \geq \epsilon$: solve $\vec{x}^{(k+1)} = \text{argmin}L_A(\vec{x};\vec{\lambda}^{(k)},\mu_k)$ unconstrained subproblem, oft same ϵ and Newton Method.
4. $\forall i(\lambda_i^{(k+1)} = \lambda_i^{(k)} + \frac{g_i(\vec{x}^{(k+1)})}{\mu_k}, \mu_{k+1} = \rho\mu_k$ for some $\rho \in (0, 1)$.
5. `return` $\vec{x}^{(f)}$. Step 2 can check $\|\nabla_{\vec{x}}L_A(\vec{x}^{(k)};\vec{\lambda}^{(k)},\mu_k)\| < \epsilon$ instead.

Corollary: $\forall$ KKT $\vec{x}^*$ with $\vec{\lambda}^*$ to non-augmented $L(\vec{x},\vec{\lambda})$, if $\vec{x}^*$ regular local min of f $\wedge H_L(\vec{x}^*,\vec{\lambda}^*)$ p.d. on $C(\vec{x}^*,\vec{\lambda}^*)$, then

1. $(\exists\gamma > 0)(\forall\mu \in (0,\gamma])(\vec{x}^*$ strict local minimiser of $L_A)$ ($\therefore \mu_k \to 0^+$ not needed, allowing better κ)
2. $(\exists\delta,\epsilon,M > 0)(\mu_k\|\vec{\lambda}^{(k)} - \vec{\lambda}^*\| \leq \delta \wedge \mu_k \in (0,\gamma) \Rightarrow \exists!$global minimiser $\vec{x}^{(k+1)}$ to $\min_{\|\vec{x}-\vec{x}^*\|\leq\epsilon} L_A(\vec{x};\vec{\lambda}^{(k)},\mu_k) \wedge \|\vec{x}^{(k+1)} - \vec{x}^*\| \leq M\mu_k\|\vec{\lambda}^{(k)} - \vec{\lambda}^*\| \wedge \vec{\lambda}^{(k)} \to \vec{\lambda}^*$ Q-linearly $\wedge\vec{x}^{(k+1)} \to \vec{x}^*$ R-linearly.

Barrier Function Method

1. Pick $\vec{x}^{(0)}$, and constraints tolerance $\epsilon > 0$. Conventionally set initial penalty $\mu_0 = 1$.
2. Crucially formulate $\min_{\vec{x}\in\mathbb{R}^n} P(\vec{x};\mu_k) := f(\vec{x}) + \mu_k B(\vec{x})$, barrier $B(\vec{x}) := \sum\phi(-g_i(\vec{x}))$ s.t. $\phi : \mathbb{R}^+ \to \mathbb{R}^+ \wedge \phi' < 0 \wedge \lim_{g_i\to 0^-}\phi(-g_i(\vec{x})) \to \infty$.
3. while $\|\nabla_{\vec{x}}P(\vec{x}^{(k)};\mu_k)\| > \epsilon$: solve $\vec{x}^{(k+1)} = \text{argmin}P(\vec{x};\mu_k)$ unconstrained subproblem, oft same ϵ, Newton Method.
4. $\mu_{k+1} = \rho\mu_k$ for some $\rho \in (0, 1)$, or dynamically set $\{\epsilon_k\}_{k\in\mathbb{N}^+}$.
5. `return` $\vec{x}^{(f)}$.

Corollaries: 1. If ($\exists$ global min $\vec{x}^*$of f)$(\forall r > 0)(N_r(\vec{x}^*) \cap X \neq \emptyset$ i.e, $\vec{x}^*$limit point of a convergent subseq$\{\vec{x_\mu}\}$), then $\lim_{\mu\to 0^+}\inf_{\vec{x}\in X}\{P(\vec{x};\mu)\} = f(\vec{x}^*) \wedge \lim_{\mu\to 0^+}\mu B(\vec{x_\mu}) = 0 \wedge$ any convergent subseq to a global min of f

2. ($\forall$convergent subseq s.t. $\{\vec{x}^{(k)}\} \to \vec{x}^*)(\vec{x}^*$ regular $\Rightarrow \vec{x}^*$ KKT with $\lambda_i^* = \lim_{k\to\infty}\mu_k\nabla\phi(-g_i(\vec{x}^{(k)}))$)

(2): Barrier function, if convergent/terminating, outputs approx-KKT-minimisers practically.