Weiestrass Theorem: S is compact $\implies f$ has a global max and a global min in S.

If f is convex, then $\{\boldsymbol{x} \colon f(\boldsymbol{x}) \le a\}$ is convex.

Epigraph E_f is convex $\iff f$ is convex.

Directional derivative at $\boldsymbol{x}$ along $\boldsymbol{d}$:

$$\nabla f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{d} = \lim_{\lambda \to 0} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda}.$$

f is convex if and only if $f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \le f(\boldsymbol{y})$.

f is a convex and continuously differentiable function, then $\boldsymbol{x}^*$ is a global minimiser $\iff$ $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*)$.

Eigenvalue Test: If $\boldsymbol{A}$ is a **symmetric real** matrix, then $\boldsymbol{A}$ is positive semidefinite $\iff$ all eigenvalues of $\boldsymbol{A}$ are non-negative.

If $\boldsymbol{A}$ is a **symmetric** matrix, then $\boldsymbol{A}$ is positive definite $\iff$ $\Delta_k < 0$ and negative definite $\iff$ $(-1)^k \Delta_k > 0$.

Taylor's Theorem: If f has continuous 2nd order partial derivatives and if the set

$$[\boldsymbol{x}, \boldsymbol{y}] \coloneqq \{\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \colon \lambda \in [0, 1]\}$$

is in the interior of D_f, then $\exists \boldsymbol{z} \in [\boldsymbol{x}, \boldsymbol{y}]$ s.t.

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^{\mathrm{T}} H_f(\boldsymbol{z})(\boldsymbol{y} - \boldsymbol{x}).$$

H_f is semidefinite $\iff$ f is convex/concave; H_f is definite $\implies$ f is strictly convex/concave; H_f is indefinite $\implies$ f is neither convex nor concave.

Coercive Function: $\lim_{\|\boldsymbol{x}\| \to \infty} f(\boldsymbol{x}) = \infty$.

$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\| \le \sqrt{2}\,\|\boldsymbol{x}\|_\infty$, where $\|\boldsymbol{x}\|_\infty = \max\{|x_i|\}$.

$\nabla f(\boldsymbol{x}^*) = 0$ and $H_f(\boldsymbol{x}^*)$ is positive definite $\implies$ $\boldsymbol{x}^*$ is a **strict** local minimiser.

If f is convex, then a local minimiser of f is a global minimiser. If f is strictly convex, then it has a unique global minimiser.

If f is convex, then any stationary point of f is a global minimiser.

$q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ is a quadratic function where $\boldsymbol{Q}$ is symmetric.

If q is defined over a convex set, then $\boldsymbol{x}^*$ is a global minimiser $\iff$ $\boldsymbol{Q}\boldsymbol{x}^* = -\boldsymbol{c}$.

Bisection Method:

$$[a_{k+1}, b_{k+1}] = \begin{cases} \left[a_k, \frac{a_k+b_k}{2}\right], & \text{if } f\left(\frac{a_k+b_k}{2}\right) f(a_k) < 0 \\ \left[\frac{a_k+b_k}{2}, b_k\right], & \text{if } f\left(\frac{a_k+b_k}{2}\right) f(a_k) > 0 \end{cases}$$

Take $x_k = \frac{a_k+b_k}{2}$. At termination, $|x^* - x_k| \le \frac{|a_k - b_k|}{2} \le \epsilon$, so we need

$$k = \left\lceil \frac{\log\left(\frac{b_1 - a_1}{\epsilon}\right)}{\log 2} \right\rceil$$

Newton's Method: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$ until $|f'(x_k)| < \epsilon$.

Multivariable Newton: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H_f(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k)$ until $\|\nabla f(\boldsymbol{x}_k)\| \le \epsilon$.

Armijo: $f(\boldsymbol{x}_0 + \alpha_k \boldsymbol{p}_0) \le f(\boldsymbol{x}_0) + \sigma\alpha_k \nabla f(\boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{p}_0$, $\sigma \in (0, 0.5)$, $\beta \in (0, 1)$.

Golden Section Method:

Set $[a_0, b_0] = [a, b]$ and take $\alpha = \frac{\sqrt{5}-1}{2}$. Compute

$$\lambda_0 = b - \alpha(b - a)$$
$$\mu_0 = a + \alpha(b - a).$$

If $f(\lambda_k) > f(\mu_k)$, then

$$a_{k+1} = \lambda_k \qquad b_{k+1} = b_k$$
$$\lambda_{k+1} = \mu_k \qquad \mu_{k+1} = \lambda_k + \alpha(b_k - \lambda_k).$$

Steepest Descent: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - t_{k-1}\nabla f(\boldsymbol{x}_k)$ until $\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$.

$(\boldsymbol{x}_{k+2} - \boldsymbol{x}_{k+1})^{\mathrm{T}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = 0.$

Conjugate Gradient: $\boldsymbol{r}_n = \boldsymbol{A}\boldsymbol{x}_n - \boldsymbol{b} = \nabla\phi(\boldsymbol{x}_n)$ where ϕ is quadratic.

$\alpha_k = -\frac{\boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{p}_k + \boldsymbol{p}_k}{\boldsymbol{p}_k^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p}_k}$.

$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$.

Lagragian: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})$.

Lagragian dual function: $\inf_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.
Dual problem: $\max_{\lambda_i \in \mathbb{R}, \mu_j \ge 0} \theta(\boldsymbol{\lambda}, \boldsymbol{\mu})$. θ is concave if finite.

Weak Duality: $f(\boldsymbol{x}^*) \ge \theta(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Equality

holds when Slater's holds ($\inf f = \sup \theta$)
Saddle point: $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Saddle point is KKT.
KKT point:

- $\nabla f(\boldsymbol{x}^*) + \sum \lambda_i \nabla g_i(\boldsymbol{x}^*) + \sum \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}$
- $g_i(\boldsymbol{x}^*) = 0$ and $h_i(\boldsymbol{x}^*) \leq 0$
- $\mu_i \geq 0$ for $i = 1, 2, \cdots, p$
- $\mu_i = 0$ for $i \notin J(\boldsymbol{x}^*)$

Complementary slackness: $\mu_i h_i(x) = 0$.
Critical Cone: $C(\boldsymbol{x}^*, \lambda, \mu) := \left\{ \boldsymbol{y} \in \mathbb{R}^n : \begin{array}{ll} \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} = 0 & \\ \nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} = 0 & \mu_j > 0 \\ \nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{y} \leq 0 & \mu_j = 0 \end{array} \right\}$
KKT 2nd necessary: $\boldsymbol{y}^{\mathrm{T}} H_L(\boldsymbol{x}^*) \boldsymbol{y} \geq 0$ for $\boldsymbol{y} \in C(\boldsymbol{x}^*, \lambda, \mu)$.
KKT sufficient: $\boldsymbol{x}^*$ is KKT and $\boldsymbol{y}^{\mathrm{T}} H_L(\boldsymbol{x}^*) \boldsymbol{y} > 0$ for nonzero for $\boldsymbol{y} \in C(\boldsymbol{x}^*, \lambda, \mu) \implies$ strict local min.
Convex constrained: convex f, convex h, linear g differentiable constraints. KKT is global min. If Slater's holds, or no h, then global min is KKT.
$\boldsymbol{x}^*$ is KKT, then $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is saddle.
$(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is KKT, then $\boldsymbol{x}^*$ optimises primal and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ optimises dual.

Subgradient:

- $\boldsymbol{\beta}(\boldsymbol{x}) = (\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x}))$: constraint vector.
- $\boldsymbol{w} = (\boldsymbol{\lambda}, \boldsymbol{\mu})$: multiplier vector.
- Lagrangian: $f(\boldsymbol{x}) + \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\beta}(\boldsymbol{x})$,
- $\boldsymbol{X}(\boldsymbol{w})$: set of $\boldsymbol{x}$ minimising $\mathcal{L}$ at $\boldsymbol{w}$.

$\boldsymbol{X}(\boldsymbol{w})$ is singleton, then θ is differentiable and $\nabla \theta(\boldsymbol{w}) = \boldsymbol{\beta}(\boldsymbol{x^*})$.
Subgradient of convex f: $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*)$, the set of which is subdifferential.
Directional derivative: $\theta(\boldsymbol{w}^*, \boldsymbol{d}) = \inf\{\boldsymbol{d}^{\mathrm{T}} \boldsymbol{\xi}(\boldsymbol{w^*})\}$.
$\partial \theta(\boldsymbol{w}) = \operatorname{conv}\{\boldsymbol{\beta}(\boldsymbol{x}) \colon \boldsymbol{x} \in \boldsymbol{X}(\boldsymbol{w})\}$ the linear span of β.

Steepest ascent direction $\boldsymbol{d}^*$: $\theta(\boldsymbol{w}^*, \boldsymbol{d}^*)$ is max among all unit vector $\boldsymbol{d}$.
$\hat{\boldsymbol{\xi}}$: Subgradient with smallest norm. $\boldsymbol{d}^* = \mathbf{0}$ if $\hat{\boldsymbol{\xi}} = \mathbf{0}$ and $\frac{\hat{\boldsymbol{\xi}}}{\|\hat{\boldsymbol{\xi}}\|}$ otherwise.

Frank-Wolfe: convex f with linear constraints.

- $z(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k)$.
- linear sol: $\hat{\boldsymbol{x}}_k$, optimal linear val: $\hat{z}_k$.
- Lower bound is the higher of the previous lower bound and $\hat{z}_k$, upper bound is $f(\boldsymbol{x}_k)$.
- $\boldsymbol{d}_k = \hat{\boldsymbol{x}}_k - \boldsymbol{x}_k$ is search direction. Find t_k s.t. $f(\boldsymbol{x}_k + t_k \boldsymbol{d}_k)$ is min.
- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k \boldsymbol{d}_k$.

Quadratic penalty: $Q(\boldsymbol{x}; \mu) = f(\boldsymbol{x}) + \frac{1}{2\mu} \sum c^2(\boldsymbol{x})$ where c are equlity constriants.

- $\nabla_{\boldsymbol{x}} Q(\boldsymbol{x}, \mu) = \nabla f(\boldsymbol{x}) + \frac{1}{\mu} \sum c(\boldsymbol{x}) \nabla c(\boldsymbol{x})$
- $H_Q(\boldsymbol{x}, \mu) = H_f(\boldsymbol{x}) + \frac{1}{\mu} \sum [\|\nabla c(\boldsymbol{x})\| + c(\boldsymbol{x}) H_c(\boldsymbol{x})] \approx H_f(\boldsymbol{x}) + \frac{1}{\mu} \sum \|\nabla c(\boldsymbol{x})\|$
- Approximate minimiser $\boldsymbol{x}_{k+1}$ of $Q(\boldsymbol{x}; \mu_k)$ e.g. with Newton with $\boldsymbol{x}_k$ as initial.
- $\mu_{k+1} = \rho \mu_k < \mu_k$ until $\|c(\boldsymbol{x}_{k+1})\| < \epsilon$.
- Limit point of $(\boldsymbol{x}_k)$ minimises f.

If $\nabla_{\boldsymbol{x}} Q(\boldsymbol{x}_k, \mu) \leq \tau_k \to 0$, $\mu_k \to 0$, $\boldsymbol{x}_k \to \boldsymbol{x}^*$ is regular, the for any subsequence $\lim \boldsymbol{x}_k = \boldsymbol{x}^*$, $\boldsymbol{x}^*$ is KKT with $\lambda_i^* = \lim \frac{c_i(\boldsymbol{x}_k)}{\mu_{k-1}}$.

Augmented Lagrangian: $L_A(\boldsymbol{x}, \boldsymbol{\lambda}, \mu) = f(\boldsymbol{x}) + \sum \lambda_i c_i(\boldsymbol{x}) + \frac{1}{2\mu} \sum c(\boldsymbol{x})^2$.

- Let $\mu_0, \tau_0 > 0$. Approximate minimiser $\boldsymbol{x}_{k+1}$ of L_A e.g. with Newton with $\boldsymbol{x}_k$ as initial.
- If $\nabla L_A(\boldsymbol{x}_{k+1}) = 0$, stop.
- $\lambda_{k+1} = \lambda_k + \frac{c(\boldsymbol{x}_{k+1})}{\mu_k}$.
- Choose new μ and τ.

Barrier function: $B(\boldsymbol{x}) = \sum \phi(c(\boldsymbol{x}))$. Commonly $\phi = -\log$.

- $P(\boldsymbol{x}; \mu_k) = f(\boldsymbol{x}) + \mu_k B(\boldsymbol{x})$.
- Approximate minimiser $\boldsymbol{x}_{k+1}$ of P s.t. $\|P(\boldsymbol{x}_{k+1}, \mu_k)\| < \tau_k$.
- Choose smaller μ and τ until $\boldsymbol{x}_{k+1}$ is regular KKT point with $\lambda = \lim \mu_k \phi'(-c(\boldsymbol{x}_k))$.