CPSC 406 Study Sheet:

Linear Least Squares:

Setting: $A: (m, n)$ $m > n$ overdetermined system, $b: (m, 1), x: (n, 1)$

Objective Function: $\min_{x} \left|\left|b - Ax\right|\right|^2 = \min_{r,x} \left\{\left|\left|r\right|\right|^2 \middle| Ax + r = b\right\}$

i. r can be treated as residuals such that $r \notin Range(A)$
ii. In general, there is no solution to the system unless $r = 0$ or $Ax = b$
iii. If $Rank(A) \neq n$, namely, A has linearly dependent columns, then the solution will not be unique (but the min value will be unique)

Solving the Linear System:

1. Normal Equations: $A^T A x_{ls} = A^T b \;\rightarrow x_{ls} = A^T A \backslash A^T b \;\; OR \;\; x_{ls} = (A^T A)^{-1} A^T b$
   i. A is assumed to be a non-singular matrix, in other words, A is assumed to be full rank.
   ii. Issues: $Cond(A^T A) = \frac{\lambda_{\max(A^T A)}}{\lambda_{\min(A^T A)}}$ might be big causing huge errors: $Precision = 16 - \log\left(Cond(A^T A)\right)$

2. Using QR decomposition:

$$A = QR = (Q_1 \quad Q_2)\begin{pmatrix} R_1 \\ 0 \end{pmatrix}, \qquad R_1 x_{ls} = Q_1^T b \;\rightarrow x_{ls} = R_1^{-1} Q_1^T b$$

   i. $Q_1: (m, n)$ is the orthogonal basis of $Range(A)$, found using the gramma-Schmidt.
   ii. $Q_2: (m, m - n)$ is the orthogonal basis of $Null(A^T)$, use gramma-Schmidt to enforce orthogonality.
   iii. $R_1 : (n, n), R: (m, n)$ are both upper-triangular matrices.
   iv. $A = Q_1 R_1$
   v. Better than normal equation because the solution has $cond(R_1)$ which has less errors. $Precision = 16 - \log\left(Cond(R_1)\right), cond(R_1) = cond(A) = \frac{\lambda_{\max(A)}}{\lambda_{\min(A)}}$

Regularized Least Squares:

Regularization: using prior knowledge (smoothness here) to regularize data:

Objective Function: $\min_{x} \frac{1}{2}\left|\left|Ax - b\right|\right|^2 - \lambda f_2(x)$

Pareto Frontier: the points below the frontier are possible & the points above are not.

Tikhonov Regularization: to promote smoothness & reduce noise.

1. Objective Function:

$$\min_{x} \left|\left|Ax - b\right|\right|^2 + \lambda\left|\left|Dx\right|\right|^2 = \left|\left|\begin{pmatrix} A \\ \sqrt{\lambda}D \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix}\right|\right|^2, \quad Dx = \begin{pmatrix} x_1 - x_2 \\ \vdots \\ x_{n-1} - x_n \end{pmatrix}, D = \begin{pmatrix} 1 & -1 & \dots \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

2. Solution to Tikhonov Regularization:

$$\nabla f(x) = 2 * \left(\begin{pmatrix} A \\ \sqrt{\lambda}D \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix}\right) * \left(\begin{pmatrix} A \\ \sqrt{\lambda}D \end{pmatrix} x\right)^T$$

$$\nabla f(x) = 2 * (A^T A x + \lambda D^T D x - A^T b)$$

$$A^T A x + \lambda D^T D x - A^T b = 0$$

$$A^T A x + \lambda D^T D x = A^T b$$

$$x = (A^T A + \lambda D^T D)^{-1} A^T b$$

Gradient

Directional Derivatives:

1. $f'(x, d) = \nabla f(x) * d$
2. $f'(x, ad) = a\nabla f(x) * d$
3. Suppose e_i are the unit directions in R^n, $f'(x, e_i) = [\nabla f(x)]_i$ is the ith element in the gradient.

Nonlinear Least Squares

1. Objective Function: $\min_{x} \frac{1}{2}\left|\left|r(x)\right|\right|^2, r(x) = Ax - b$

2. Using the Gauss-Newton Method to solve:
   i. Start from an arbitrary point: x_0
   ii. For $k = 0,1,2 \dots, n$:
      a. Linearize the residuals at kth iteration: $r^k(x) = r(x_k) + J(x_k)(x - x_k) + o\left(\left|\left|x - x_k\right|\right|\right)$
      b. Find x_{k+1} by minimising the linearized residuals:

	$x_{k+1} = argmin_x \left(\frac{1}{2}\left\|\left\|r^k(x)\right\|\right\|^2\right) = argmin_x \left(\frac{1}{2}\left\|\left\|r(x_k) + J(x_k)(x - x_k)\right\|\right\|^2\right)$ $J(x_k)x = r(x_k) - J(x_k)x_k$ $x_{k+1} = J(x_k)\backslash(r(x_k) - J(x_k)x_k), J(x_k) = \nabla r(x_k)^T$ $OR \quad x_{k+1} = A_k\backslash b_k, \; A_k = J(x_k), b_k = r(x_k) - J(x_k)x_k$ c. If $\left\|\left\|x_{k+1} - x_k\right\|\right\|^2 < \varepsilon$, exit the loop. iii. Return x_{k+1}. 3. Optimality: x^* is a local min if i. Necessary conditions: 1. $\nabla f(x^*) = 0$, 2. $\nabla_2 f(x^*) \geq 0$ (Hessian is semi-positive definite) ii. Sufficient conditions: 1. $\nabla f(x^*) = 0$, 2. $\nabla_2 f(x^*) > 0$ (Hessian is positive definite) 4. Coerciveness & Unboundedness: i. Coerciveness: $\lim_{\|\|x\|\| \to \infty} f(x) = \infty$, this implies that there is a minimum. ii. Unboundedness: $\lim_{\|\|x\|\| \to \infty} f(x) = -\infty$, this implies that there is no minimum.
Gradient Descent:	Step Size Selection: i. Constant Step Size a. Using Lipschitz Community of Gradient: $\left\|\left\|\nabla f(x) - \nabla f(y)\right\|\right\| < L\left\|\left\|x - y\right\|\right\|$ b. $L = \left\|\left\|A\right\|\right\| = \lambda_{MAX}(A)$ c. $\alpha \in \left[0, \frac{2}{L} = \frac{2}{\lambda_{MAX}(A)}\right]$ ii. Exact Line Search a. May not always work but always work for quadratic functions. b. $\alpha = \frac{-\nabla f(x)^T d}{d^T A d}$ iii. Backtracking Line Search: a. Choose α such that $f(x_k) - f(x_k + \alpha d_k) \geq -\mu\alpha\nabla f(x_k)d_k$ Algorithm: While $(x_k) - f(x_k + \alpha d_k) < -\mu\alpha\nabla f(x_k)d_k$: $\alpha = \beta\alpha$ Search Directions: $\nabla f(x, d) < 0$, Steepest Descent: $d = -\nabla f(x, d)$ Descent Algorithm: 1. start from x_0 2. For $k = 0, 1, \dots, n$: i. Compute search direction d_k ii. Choose Step size α_k using one of the three methods above. iii. $x_{k+1} = x_k + \alpha d_k = x_k - \alpha\nabla f(x_k)$ (If using steepest descent). iv. Exit when $\left\|\left\|\nabla f(x_{k+1})\right\|\right\| < \varepsilon$ 3. Return x_{k+1}. Newton's Method: 1. start from x_0 2. For $k = 0, 1, \dots, n$: i. Compute search direction d_k, $\nabla_2 f(x)d = -f(x_k)$ (using Newton's direction, convergence require $\nabla_2 f(x) > 0$ ii. Choose Step size α_k using one of the three methods above. iii. $x_{k+1} = x_k + \alpha d_k$ iv. Exit when $\left\|\left\|\nabla f(x_{k+1})\right\|\right\| < \varepsilon$ 3. Return x_{k+1}. Scaled Descent: 1. start from x_0 2. For $k = 0, 1, \dots, n$: i. Choose Step size α_k using one of the three methods above.

	ii. $x_{k+1} = x_k - \alpha D \nabla f(x_k),\ D = SS^T,\ S = \begin{pmatrix} S_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & S_n \end{pmatrix}, x = sy$ iii. Exit when $\lVert \nabla f(x_{k+1}) \rVert < \varepsilon$ 3. Return x_{k+1}.
Positive definiteness & Cholesky Factorization	Positive definite matrix: Matrix $A: (n, n)$ is positive definite if for all column vectors x: $x^T A x > 0$ Properties of Positive Definite Matrices: 1. For any full rank matrix X, if A is positive definite, then $X^T A X$ is also positive definite. 2. If a matrix has all positive eigenvalues $\leftrightarrow$ the matrix is positive definite. 3. A is positive definite $\leftrightarrow$ A has Cholesky Factorization. 4. A is positive definite $\rightarrow$ the first entry a_{11} of A is positive: $a_{11} > 0$ Cholesky Factorization: 1. $A = \begin{pmatrix} a_{11} & w^T \\ w & k \end{pmatrix}$, a_{11} is the first top left entry of A, w is the remainder of the first column. 2. $A = \begin{pmatrix} \alpha & 0 \\ w/\alpha & I \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & k - \frac{ww^T}{a_{11}} \end{pmatrix} \begin{pmatrix} \alpha & w^T/\alpha \\ 0 & I \end{pmatrix}$, $k - \frac{ww^T}{a_{11}}$ is also positive definite.
Linear Constraint	Setting: $A: (m, n), m < n$: underdetermined system, fewer equations than variables Objective Function: $\min_{x \in R^n} f(x)\ \ subj\ to\ Ax = b$ Feasible Sets: $F = \{x \mid Ax = b\} = \{\bar{x} + zp \mid p \in R^{n-m}\}$ i. $\bar{x}$ is a particular solution such that $A\bar{x} = b$, $Az = 0, < A^T, z > = 0\ \rightarrow z \in Null(A)$ ii. $Range(z) = Null(A)\ \perp Range(A^T),\ Null(z^T) = \ Range(A^T)$ Reduced Problem: $\min_{p \in R^{n-m}} f(\bar{x} + zp)$, suppose p^* is the optimal solution, $x^* = \bar{x} + zp^*$: i. Optimality condition: $\nabla_p f(\bar{x} + zp^*) = 0 \rightarrow 0 = z^T \nabla f(x^*) \rightarrow 0 =< z, x^* > \rightarrow x^* \in Null(z^T) = Range(A^T)$ First-Order Necessary Conditions: A point x^* is a local min of $\min_{x \in R^n} f(x)\ \ subj\ to\ Ax = b$ only if there exists a vector $y \in R^n$: i. Optimality: $\nabla f(x^*) = A^T y\ \ \leftrightarrow \nabla f(x^*) \in Range(A^T)\ \ \leftrightarrow z^T \nabla f(x^*) = 0\ \leftrightarrow \nabla f(x^*)^T p = 0, \forall p \in Null(z)$ ii. Feasibility: $Ax^* = b$ Second-order Optimality: 1. Second-order necessary conditions: i. Optimality: $\nabla f(x^*) = A^T y\ \ \leftrightarrow \nabla f(x^*) \in Range(A^T)\ \ \leftrightarrow z^T \nabla f(x^*) = 0\ \leftrightarrow \nabla f(x^*)^T p = 0, \forall p \in Null(z)$ ii. Feasibility: $Ax^* = b$ iii. $z^T \nabla_2 f(x^*) z \geq 0\ \leftrightarrow p^T \nabla_2 f(x^*) p \geq 0, \forall p \in Null(z)$ 2. Sufficient conditions: all the same except for: $z^T \nabla_2 f(x^*) z > 0\ \leftrightarrow p^T \nabla_2 f(x^*) p > 0, \forall p \in Null(z)$. Conditions changed from semi-positive definite to strictly positive definite.
Convexity	Convex Sets: A set $C \in R^n$ is convex if for any points $x, y \in C$ and any $\lambda \in [0,1]$, $\lambda x + (1 - \lambda) y \in C$ Examples of Convex Sets: 1. Affine/line is convex: $L = \{z + td \mid t \in R\}$, $z \in R^n, d \neq 0 \in R^n$ 2. Hyperplane: $H_{\alpha,\beta} = \{x \in R^n \mid \alpha^T x = \beta\}$, $\alpha \in R^n \neq 0, \beta \in R$ 3. Half-space: $H^-_{\alpha,\beta} = \{x \in R^n \mid \alpha^T x \leq \beta\}$, $\alpha \in R^n \neq 0, \beta \in R$ 4. Norm Ball: $\{x \in R^n \mid \lVert x - c \rVert \leq r\}$, $c \in R^n$ is the center, $r \in R$ is the radius. 5. Convex Hulls of a set: $Conv(s) = \{\sum_{i=1}^{k} \lambda_i x_i \mid x_i \in S,\ \sum_i \lambda_i = 1, k \in N\}$

	6. Simplex: $\Delta_n = \{x \in R^n \mid \sum_{i=1}^{n} x_i \leq 1, k \in N\}$ 7. Unit Simplex: $\overline{\Delta_n} = \{x \in R^n \mid \sum_{i=1}^{n} x_i = 1, k \in N\}$ Set Operations that Preserve Convexity: 1. intersections of convex sets are convex 2. Additions of Convex sets are convex 3. Image: If $C \in R^n$ is a convex set & matrix $A: (m, n)$, then $A(c) = \{Ax \mid x \in C\}$ is also convex. Convex Functions: A function: $f: C \in R^n \rightarrow R$, C is convex, is convex if: $$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$ Strict convex: $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$ Common Convex Functions: 1. Affine/line function: $a^T x + \beta$ for some $\alpha \in R^n, \beta \in R$ (convex & concave) 2. Exponential: e^{ax} for any $a \in R$ 3. Powers: x^a on R_{++} for all $a \geq 1$ or $a \leq 0$, Note that when $0 < a < 1$ the function is concave 4. absolute value: $\|x\|^p, p \geq 1$ 5. Negative entropy: $xlog(x)$ when $x \in R_{++}$ 6. Norms: all the norms are convex; proved using the triangular inequality: $\|\|x+y\|\| \leq \|\|x\|\| + \|\|y\|\|$ Function Operations that Preserve Convexity: 1. Non-negative multiplies. 2. Sum of the convex functions are also convex 3. Composition of convex functions with an affine function is also convex: i.e. $e^{a^T x + \beta}$ Theorems to Prove Convexity of Functions: 1. Let $f: C \rightarrow R$ be continuously differentiable over $C \in R^n$, then f is convex iff: $$f(x) + \nabla f(x)(z - x) \leq f(z), \forall x, z \in C$$ 2. . Let $f: C \rightarrow R$ be continuously twice differentiable over $C \in R^n$, then f is convex iff: $$\nabla_2 f(x) \geq 0 \rightarrow Pos.def$$ Convex Optimality: 1. For a convex function, if x^* is a local min, then it's a global min. 2. Unconstraint case: Optimal $\rightarrow$ $\nabla f(x^*) = 0$ 3. Constraint case: $\nabla f(x^*)^T (x - x^*) \geq 0, x \in C$ (all feasible directions are non-decreasing). i. In other words: $-\nabla f(x^*) \in N_c(x^*), N_c(x^*) = \{g \in R^n \mid g^T(z - x) \leq 0, z \in C\}$ ii. $\leftrightarrow$ $-\nabla f(x^*) \in Range(A^T)$
Projection	Projection Theorem: $proj_c(x) = g(z) = \min_{z \in C} \frac{1}{2} \|\|z - x\|\|^2$ 1. If the objective function is convex then projection is unique. 2. If $x \in C$, then $proj_c(x) = x$ 3. $-\nabla g(z) = -(z - x) = x - z \in N_C(z), N_c(z) = \{g \in R^n \mid g^T(c - z) \leq 0, z \in C\}$ Projected Gradient Method: 1. start from x_0 2. For $k = 0, 1, \ldots, n$: i. Choose Step size α_k using one of the three methods above. ii. $x_{k+1} = proj_c(x_k - \alpha_k * \nabla f(x_k))$ iii. Exit when $\|\|\mathrm{x_k} - x_{k+1}\|\| < \varepsilon$ 3. Return x_{k+1}. Stationarity of Projected Gradient: $x^* \in argmin_{x \in C} f(x)$ with C closed & convex, the $f: R^n \rightarrow R$ is convex differentiable if and only if: $$x^* = proj_c(x^* - \alpha_k * \nabla f(x^*))$$

Convergence of Gradient Descent:	
Linear Programming	Geometry of Linear Programming: Suppose we have a polyhedron: $P = \{x \mid Ax \leq b\}$, $A = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix}$, $A: (m, n)\ m < n$. Polyhedrons are convex because they are intersections of half-spaces. 1. Equivalent Concepts: i. Extreme points: $x \in P$ is an extreme point of P if there does not exist two vectors $y, z \in P$ such that: $$x = \lambda y + (1 - \lambda) z$$ ii. $x \in P$ is a vertex of P if there exists aa vector $c \neq 0$ such that: $$c^T x < c^T y, \qquad \forall y \in p, y \neq x$$ iii. x^* is a basic solution if $a_{i,1}, \ldots, a_{i,n}$ are linearly independent $\leftrightarrow Bx^* = b_B \leftrightarrow Rank(B) = n$ Settings: $A = (B \quad N)$ where $B: (m, m)$ is a basis of A, let the indices in B be $\beta = \{i_1, .., i_k\}$, $b_B = \begin{pmatrix} b_{i_1} \\ \vdots \\ b_{i_k} \end{pmatrix}$ a. Active Constraints: $a_i^T x^* = b_i, i \in \beta$ b. Inactive feasible constrains: $a_i^T < b_i, i \in N$ c. Inactive infeasible constraints: $a_i^T x^* > b_i, i \notin B \cup N$ Properties of Polyhedron: 1. P contains a full line $\leftrightarrow$ P has no extreme points 2. P is unbounded $\leftrightarrow$ P contains a half-line. Converting Generic Polyhedron to Standard Form: Suppose we have a generic polyhedron $P = \{x \mid Ax = b,\ Cx \leq d\}$ & we want to convert it to standard form: $p = \{x \mid Ax = b, x \geq 0\}, b \geq 0$, the following steps should be taken: 1. Ensure that all converted $b_i \in b$ are positive: i. For any $b_i < 0$ in generic form $Ax = b$, replace $a_i x_i = b_i \ \rightarrow (-a_i)x = (-b_i)$ ii. For $d_i < 0$, replace $c_i^T x \leq d_i \ \rightarrow -c_i^T x \geq -d_i,\ c_i^T x \geq d_i \ \rightarrow -c_i^T x \leq\ -d_i$ 2. Converting Free Variables x_i (x_i has no constraints): $$x_i = x_i' - x_i'', x_i' \geq 0 \ \&\ x_i'' \geq 0$$ i. x_i' encodes positive part of x_i, x_i'' encodes the negative part of x_i ii. Optimal solution must have $x_i' * x_i'' = 0$ 3. Using Slack & Surplus to Convert Inequality Constraints: i. Replace $c_i^T x \leq d_i \ \rightarrow c_i^T x + s_i = d_i, s_i \geq 0$ ii. Replace $c_i^T X \geq d_i \ \rightarrow c_i^T x - s_i = d_i, s_i \geq 0$ Basic Solution in Standard Form: 1. Setting: n variables, $m + n$ constraints (m equality constraints from $Ax = b$, n inequality constraints $x \geq 0$) 2. $\bar{A}x = \begin{pmatrix} B & N \\ 0 & I \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}, x_N = 0, Bx_B = b, B: (m, m), A = (n, n), N = (n - m, n - m)$ 3. Solution is degenerate if some elements in x_B is also 0.

<table>
<tr><td></td><td>
Simplex Method:
Assumptions: problem in LP standard Form:

$$\min_x c^T x \ subj\ Ax = b, x \geq 0$$

Simplex Method:
Start with basic variables with indices $B = \{\beta_1, \dots, \beta_m\}$ ($x_{i\in B}$ are nonzero) and non-basic variables $N = \{n_1, .. n_{n-m}\}$ such that $A = [B, N], c = \binom{c_B}{c_N}$

1. For each iteration: $k = 0, \dots, n$:
   i. Compute $Bx_B = b \ \rightarrow x_B$
   ii. Compute y such that $B^T y = c_B \ \rightarrow y$
   iii. Calculate reduced cost: $z_N = c_N - N^T y \ \rightarrow Z_n,$ choose the most negative entry n_k in Z_n to enter the basis.
   iv. Solve $Bd_B = -a_{n_k}$
   v. Kick the basic variable with index: $q = argmin_{q\in\{1,\dots,m\}\&\, d_{\beta_q}<0} - \frac{x_{\beta_q}}{d_{\beta_q}}$
   vi. Form the new basic & non-basic basis & start over again.

Optimality: No improving directions exists for each $j = 1, \dots, n$ such that

$$x_j = 0 \,\&\, z_j \geq 0 \ \ OR \ \ x_j \geq 0 \,\&\, z_j = 0 \text{ (Must hold for basics)}$$
</td></tr>
<tr><td>Duality</td><td>
Primal Problem:

$$\min_x c^T x \ subj\ to\ Ax = b, x \geq 0$$

i. n variables, m constraints (because A is $m \times n, n > m$, underdetermined)
ii. Suppose optimal x^*, then the unique optimal value would be $p^* = c^T x^*$

Dual of the Linear Problem:

$$\max_y b^T y \ subj\ to\ c - A^T y \geq 0$$

In standard form:

$$\max_{y,z} b^T y \ subj\ to\ A^T y + z = c, z \geq 0$$

i. We applied “slack & surplus method” to obtain the value of z such that $z = c - A^T y$

Derivation of the Dual of the Linear Problem:
1. Consider the relaxed version of the primal problem (by converting the constraint to a price of violation)

Relaxed Problem: $\min_x c^T x + y^T(b - Ax) \ \ subj\ to\ x \geq 0,\ y$: price of violating constraints

i. Relaxed problem provides a lower bound for optimal value p^* (by definition)
ii. Dimensions: $A = (m, n), x = (n, 1), b = (m, 1), y = (m, 1)$
iii. $g(y) = \min_{x\geq 0}\{c^T x + y^T(b - Ax)\} \leq c^T x^* + y^T(b - Ax^*) = c^T x^* = p^*$
iv. $y^T(b - Ax^*)$ cancelled out because $Ax^* = b$ by constraint.

2. Then by simplifying $g(y)$ as a function of y, we obtain:

$$\begin{aligned} g(y) &= \min_{x\geq 0}\{c^T x + y^T(b - Ax)\} \\ &= \min_{x\geq 0}\{c^T x + y^T b - y^T Ax\} \\ &= y^T b + \min_{x\geq 0}\{c^T x - y^T Ax\} \\ &= b^T y + \min_{x\geq 0}\{x^T(c - A^T y)\}, \text{ dimension of two parts here are } (1,1) \\ &= b^T y + \min_{x\geq 0}\{x^T(c - A^T y)\} \\ &= \begin{cases} b^T y & if\ c - A^T y \geq 0 \\ -\infty & otherwise \end{cases} \end{aligned}$$

If $c - A^T y < 0$, then I can choose x arbitrarily such that $g(y) \rightarrow -\infty$

Weak Duality:
Suppose x is primal feasible (constraints for primal problem satisfied: $Ax = b, x \geq 0$) and (y, z) is dual feasible (constraints for dual problem satisfied: $A^T y + z = c, z \geq 0$), then the primal objective is bounded below by the dual objective:

$$c^T x = (A^T y + z)x = y^T Ax + z^T x = y^T b + z^T x \geq y^T b, \ \ z \geq 0$$

Weak Duality Theorem:
If (x, y, z) is primal/dual feasible, then for value p:
</td></tr>
</table>

<table>
<tr>
<td></td>
<td>
i. the primal value is an upper bound for the dual value.
ii. the dual value is a lower bound for the primal value.

Complementarity:
The bound is tight (primal value = dual value) when x & y are complementary, namely $x^T z = 0$:
$x_j = 0 \ \& \ z_j \geq 0 \ \ or \ \ x_j \geq 0 \ \& \ z_j = 0$

Optimal Conditions:
1. Simplex maintains primal feasibility at every iteration: $Ax = b, x \geq 0$
2. Method defines y via $B^T y = c_B, \ z = c - A^T y$ & maintains complementarity: $x_B \geq 0$ & $z_B = 0$ and $x_N = 0$ & $z_N >=< 0$.
3. Exit when $z \geq 0$ such that (y, z) is dual feasible: $A^T y + z = c, z \geq 0$

Strong Duality Theorem: if an LP has an optimal solution, so does it dual, and then optimal values for dual & primal problems are equal.

Theorem: the primal-dual triple (x, y, z) is optimal iff
1. Primal Feasible: $Ax = b, x \geq 0$
2. Dual Feasible: $A^T y + z = c, z \geq 0$
3. Complementarity: $x^T z = 0$
</td>
</tr>
<tr>
<td>Matrix Game:</td>
<td>
Let matrix A denotes to the amounts that Y pays X such that a_{ij} represents the specific amount by X taking strategy j & Y taking strategy i.
$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{nm} \end{pmatrix}, A = (m, n)$
X strategy: Choose x subject to $e^T x = 1 \ OR \ \sum_{j=1}^{n} x_j = 1$
Y strategy: Choose y subject to $e^T y = 1 \ OR \ \sum_{i=1}^{m} y_i = 1$
Total Expected Payoff: $y^T Ax = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j y_i$

Player Y's analysis:
Suppose Y chooses y as his strategy, then X will best defend by choosing x to maximise the expected payoff:
$\max_x y^T Ax \ \ subj \ to \ e^T x = 1, e^T y = 1, y \geq 0, x \geq 0$
Then, Y should choose y to protect against the worst possible case when X knows what Y will do: given y,
$\min_y (\max_x y^T Ax) \ \leftrightarrow \ \max_j (y^T A)_j$
1. Solving for Y's strategy:
i. From LP theory, a basic optimal solution exists implies that x^* only has 1 nonzero component = 1. (Because there is only 1 equality constraint for inner problem x)
ii. Original Problem: $\min_y (\max_x y^T Ax) \ \ subj \ to \ e^T y = 1, y \geq 0$
iii. Reformulate as an LP: $\min_{y,v} v \ \ subj \ to \ ve \geq A^T y, e^T y = 1, y \geq 0$
Player X's strategy: $\max_x \left(\min_y y^T Ax \right) \ subj \ to \ e^T x = 1, e^T y = 1, y \geq 0, x \geq 0$
i. Similarly, y^* only has 1 nonzero component (=1)
ii. Reformulated problem: $\max_{x,\lambda} \lambda \ \ subj \ to \ \lambda e \leq Ax, e^T x = 1, x \geq 0$

MiniMax Theorem: $\lambda = v$, namely X's worst-case expected win = Y's worst-case expected loss.
i. X & Y analysis are dual pairs such that their optimal values should be the same (by strong duality). X's the primal problem & Y is the dual problem.
</td>
</tr>
</table>