Deterministic OR: an informal review (basic courses)

A. Bochkarev; abochka@clemson.edu
ver. 2019-1.3

1 General notes

- **Optima**
  - global min: $\forall x \in S:\ f(x) \ge f(\bar x)$
  - local min: $\exists \varepsilon > 0:\ \forall x \in N_\varepsilon(\bar x) \cap S:\ f(x) \ge f(\bar x)$
    - strong: ∃ a neighborhood s.t. it is the only local min in it
    - weak: not strong (surprisingly)
  - Notes: Strong min → Strict min; Strict min, though, does not necessarily implies strong min
- **Other important points**
  - **Stationary point** - s.t. $\nabla f = o$
  - **inflection point** -- roughly, when a curve changes between concave <-> convex (or where a function has a convex and concave cross sections)
- **Optimization problem -- "general" form** (I.e. used throughout this map, unless stated otherwise): $\min f(x)$ s.t. $g_i(x) \le 0,\ i=\overline{1,m}$; $h_i(x)=0,\ i=\overline{1,p}$
- **Optimization problem -- "convex"**: $f(x),\ g_i(x), i=\overline{0,m}$ – convex; $h_i(x), i=\overline{1,p}$ – affine
- **Other** – Wording: Now, this is very loose, but: - **strict** usually means we enforce strict inequality somewhere; (forbid equality); - **strong** usually means there is not just inequality, but some actual "gap" between something (forbid these infinitesimal, vanishing differences)

2 Linear algebra / useful stuff

2.1 On matrices
- Rank A = # of linearly independent rows / cols (the smallest of)
- A is full-rank -- iff rank A = min(# rows, # cols)
- Rank doesn't change if we transpose A
- ***Def.* Range** = column space of A (Note that Ax = linear combination of cols of A with x_i's)
- Solutions to systems of linear equations: $Ax=b$
  - Obviously, exists if *b* is in *Range of A*
  - Whether the solution exists sometimes can be efficiently checked with Farkas' lemma & co.
  - Solution *methods*: Gauss – $O(n^3)$; Iterative refinement: Guess initial solution; solve eq 1 to update x1; solve eq 2 to update x2 and so on; Faster, but does not always work
  - Some methods can achieve significant imrpovements in speed by exploring special structure of the matrix (if it is there) (Sparsity, Hessenberg systems, etc.)

2.2 Norms
- **Definition:** p is a norm iff: (roughly speaking) $p: S \to \mathbb{R}$; $\forall a,b \in S, \lambda \in \mathbb{R}$: $p(a+b) \le p(a)+p(b)$; $p(\lambda a) = |\lambda| p(a)$; $p(a)=0 \to a=o$
- Examples (of norms over vectors): $\|x\|_1 = \sum |x_i|$; $\|x\|_2 = \sqrt{\sum x_i^2}$; $\|x\|_p = (\sum |x_i|^p)^{\frac{1}{p}}$; $\|x\|_\infty = \max_i |x_i|$
- Cauchy-Bunyakovsky-Schwarz: $x^T y \le \|x\|_2 \times \|y\|_2$

Key object – linear combination $\sum \lambda_i x_i$ (sum of x's weighted by lambda's)
- Just "linear": Arbitrary λ's; If ∃ non-trivial linear combination =0 → vectors are "lin/dep"; Ax = 0 has non-triv sol iff rank A < n
- Conic: $\lambda_i \ge 0$; A cone: C is a cone iff: $\forall x \in C, \lambda \ge 0:\ \lambda x \in C$; Not necessarily convex ("Ice-cream" vs. "V" for Vendetta)
- Affine: $\sum \lambda_i = 1$
- Convex (=Conic affine comb): $\sum \lambda_i = 1$, and $\lambda_i \ge 0$
- Convex ⊂ Conic ⊂ Linear
- Convex ⊂ Affine ⊂ Linear
- **Hulls**: $\{\sum \lambda_i x_i | x_i \in S$ (and < condition >)$\}$: Conv(S) (The smallest convex set that contains S - "convex hull"); Aff(S) ("affine hull"); Cone(S); Span(S) (linear combinations)

2.3 Useful notes from calculus
- Thm (Weierstrass): A continuous function over a (non-empty) compact set does attain its minimum and maximum (recall: compact = bounded and closed)
- Useful facts about multi-variate "derivatives"
  - Definitions: $[\nabla f]_i = \left[\frac{\partial f}{\partial x_i}\right]$ – column vector; $[Df]_{ij} = \left[\frac{\partial f_i}{\partial x_j}\right]$ - row vector; $[H[f]]_{ij} = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] = D[D^T f]$
  - "Taking derivatives" (IMO, a good short note on the matter: https://atmos.washington.edu/~dennis/MatrixCalculus.pdf)
    - "Chain rule": $f(x) = g(h(x))$; $D[f] = Dg(h(x)) \cdot D[h(x)]$
    - Selected expressions: $\nabla x^T a = \nabla a^T x = a$; $\nabla(x^T A x) = (A^T + A)x$; $D[a^T b] = b^T Da + a^T Db$; $D[a^T a] = 2a^T Da$; $D[Ax] = A$; $D[Ag(x)] = AD[g(x)]$; $D[y^T A x] = y^T A$; $D[a^T A b] = b^T A^T Da + a^T A Db$
- Differentiability
  - f - "differentiable" at $\bar x \in \text{int} S$ if: $\exists \nabla f(\bar x),\ \alpha: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}:\ \lim_{x \to \bar x} \alpha(\bar x; x - \bar x) = 0$ and: $\forall x \in \text{int} S:\ f(x) = f(\bar x) + \nabla^T f(\bar x)(x-\bar x) + \|x - \bar x\| \cdot \alpha(\bar x; x-\bar x)$
    - **Note:** existence of gradient does not imply, generally speaking, differentiability
      - For **one variable** case, existence of the derivative at a point is sufficient for differentiability at that point
      - For **multi-variable** case -- sufficient condition: $\exists \varepsilon > 0: \forall x \in N_\varepsilon(\bar x) \exists \nabla f(x)$; The gradient is continuous at this point
  - f - "twice differentiable" at $\bar x \in \text{int} S$ if: $\exists \nabla f(\bar x),\ Hf(\bar x) \in \mathbb{R}^{n \times n}$ - symmetric, $\alpha: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}:\ \lim_{x\to\bar x} \alpha(\bar x; x-\bar x) = 0$ and: $\forall x \in \text{int} S:\ f(x) = f(\bar x) + \nabla^T f(\bar x)(x-\bar x) + \frac12 (x-\bar x)^T Hf(\bar x)(x - \bar x) + \|x-\bar x\|^2 \cdot \alpha(\bar x; x - \bar x)$
- "Analytic" function: the one that is locally given by a convergent series (i.e. Taylor expansion exactly "works" in some open neighborhood); **Note:** analytical functions are infinitely differentiable ("smooth"), but smooth ones are not necessarily analytical.
- **Thm.** "Trust the gradient": $f: \mathbb{R}^n \to \mathbb{R}$ – differentiable at $\bar x$. Then: $\nabla f(\bar x) d < 0 \to \exists \delta > 0:\ \forall \lambda \in (0,\delta):\ f(\bar x + \lambda d) < f(\bar x)$; **Corollary.** For differentiable functions (over an open set): $\bar x$ – local min → $\nabla f(\bar x) = 0$
- Mean value theorem
  - MVT ("traditional"): $f: S \to \mathbb{R}$, differentiable, S – an open ball in $\mathbb{R}^n$; $\forall x \in S, \exists \bar x \in S:\ f(x) = f(\bar x) + \nabla^T f(\bar x)(x - \bar x)$
  - Taylor's thm (with remainder): $f: S \to \mathbb{R}$, twice differentiable, S – a ball in $\mathbb{R}^n$; $\forall x \in S, h \in \mathbb{R}^n: (x+h) \in S\ \exists c \in (0,1):\ f(x+h) = f(x) + \nabla^T f(x) h + \frac12 h^T Hf(x+ch)h$

2.4 On quadratic forms
- A – "positive-(semi)definite" iff: $\forall x \in \mathbb{R}^n\ \ x^T A x \ge 0$
- Fun fact: A – PSD and $x^T A x = 0 \to Ax = o$
- Quadratic problems (unconstrained): $f(x) = \frac12 x^T A x + b^T x + c$
  - **minimal value -- f***: $-\frac12 \bar x^T A \bar x + c$: if $-b \in \mathcal{R}(A)$; $\bar x$ is a solution to $Ax+b=o$; $-\infty$ otherwise
  - Note no optimum if $\frac12(A + A^T) \not\succeq 0$
  - "Derivatives": $\nabla f = \frac12 (A + A^T)x + b$; $\nabla^2 f = \frac12 (A+A^T)$
- PSD quadratic form (we are discussing symmetric matrices): All eigenvalues are nonnegative; All leading principal minors are positive -> PD (Sylvester's criterion, LPM is a determinant of an upper-left square submatrix); for PSD all PMs need to be nonnegative
- Sum/conic combination of (s)PSD matrices is, obviously, (s)PSD

2.5 Selected matrix decompositions
- Diagonalization: $\exists P, D:\ A = P^{-1} D P$; entries of D – eigenvals of H; A is diagonalizable if it has n distinct eigenvalues (sufficient, not necessary); Real symmetric matrices are diagonalizabe (by orthogonal matrices)
- Note: "square root" of a matrix: Sometimes it is handy to decompose A into a product (a "square"): $A = QQ^T$; e.g., for a PSD symmetric matrix such Q can be easily found by diagonalization
- Cholesky decomposition: Decomposition of a Hermitian, PD matrix into a lower-triangular and its conjugate transpose
  - Hermitian matrix: such that $A = \overline{A}^T (=A^H)$; Obviously, every **real symmetric matrix** is Hermitian
  - Conjugate transpose (aka Hermitian transpose): $[A^H]_{ij} = \overline{A_{ji}}$; take the matrix A; transpose it; conjugate every entry (i.e. invert the sign of the imaginary part)
  - PD -- see above
  - Lower triangular: i.e. everything above the main diagonal is zero
  - $A = LL^H$; Every Hermitian PD matrix has a unique Cholesky decomposition (thus, every real PD matrix!)
  - $A = LDL^H$: where D is a diagonal matrix; L is unitriangular (i.e. triangular with entries on the main diagonal being 1)

3 Convex sets

3.1 Definitions
- **Def** $\forall x, y \in S, \lambda \in [0,1]:\ \lambda x + (1-\lambda) y \in S$
- **Alternative** $(\forall k > 0:\ \forall \lambda_i \ge 0, \sum \lambda_i = 1:\ x_1, \ldots, x_k \in S \to \sum \lambda_i x_i \in S)$

3.2 Convex hull
- $\text{conv}(S) = \{x:\ x = \lambda x_1 + (1-\lambda) x_2,\ x_1, x_2 \in S, \lambda \in (0,1)\}$
- **Some properties**: is a convex set, obviously; is the "smallest" convex set containing S (well, in a sense that it is the intersection of all convex sets containing S)

3.3 Representation of a point (Caratheodory thm)
Any point in a convex set can be represented as a convex combination of at most (n+1) points: $S \subseteq \mathbb{R}^n$ – convex then $\forall x \in S:\ \exists p_1, \ldots, p_{n+1} \in S, \lambda_1, \ldots, \lambda_{n+1} \in \mathbb{R}_+:\ \sum \lambda_i = 1$ and $x = \sum_{i=1}^{n+1} \lambda_i p_i$

Proof idea: x is in a conv(S), as S is convex; then we can represent it as a convex combination of k points (where k is some integer); if k<=n+1 -- we are done; if k>n+1: note that (x2-x1, x3-x1, ...) -- are linearly dependent in R^n; So after some thought I can find a non-trivial LC of x's =0 -- and throw away one x_j from the initial representation of x; Repeat until k=n+1

3.4 Facts about sets convexity
- Image of a convex set under an affine f is convex (Say, Ellipsoid is convex as it is an image of a unit ball under an affine transformation)
- Inverse image of a convex set under affine f is convex (i.e. f(x) = Ax + b)
- **Note.** Any convex set can be represented as an intersection of infinite number of half-spaces (hint: supporting hyperplanes)
- S, T – convex $\to S \times T$ is convex
- Level set S_α: $S \ne \emptyset$ – convex, dom $f = S$, $\alpha \in \mathbb{R}$; $S_\alpha = \{x \in S | f(x) \le \alpha\}$; if f is convex → $\forall \alpha \in \mathbb{R}:\ S_\alpha$ is convex; the converse is not true (though, level sets are so cool that you might want to define a separate class of functions...)
- **Standard simplex**: $\{x \in \mathbb{R}^n | \sum x_i = 1, x \ge 0\}$
- **Thm.** A set is convex **iff** it's intersection with any line is convex
- $S = \cap_{\alpha \in A} S_\alpha$ – convex if all S_α are convex

4 A splash of topology
- Neighborhood: **Def**: $N_\varepsilon(x) = \{x: \|x - \bar x\| < \varepsilon\}$; **Def**: $\dot N_\varepsilon(x) = \{x \ne \bar x:\ \|x - \bar x\| < \varepsilon\}$
- Open set: **Def** $\forall x \in S\ \exists N_\varepsilon(x) \subseteq S$; or: such that $S = \text{int}(S)$
- Boundary: Boundary point of S: **Def** $\bar x$ is b.p. $\iff \forall \varepsilon > 0:\ \dot N_\varepsilon(\bar x) \cap S \ne \emptyset,\ \dot N_\varepsilon(\bar x) \cap S^C \ne \emptyset$; Boundary (set) of S: **Def** ∂S = set of all boundary points
- Closed set, lim points and boundary: Limit point: **Def**: $\bar x:\ \exists\{x_n\}: \forall n\ x_n \in S$ and $\lim_{n\to\infty} x_n = \bar x$; Closed set: the one that contains all its l.p.; Or: such that cl(S) = S; Closure of a set: $\text{cl}(S) = S \cup \{\text{l.p. of } S\}$; Or: closure point – a point, any N_ε of which contains at least one pt from S
- Interior: $\text{int}(S) = S \setminus \partial S$; int. point – if some ball centered at it lies completely in S; is an open subset of S; is a union of all open subsets of S -- the "largest" open subset of S; Relative interior: $\text{relint} S = \{x \in S:\ \exists \varepsilon > 0, N_\varepsilon(x) \cap \text{aff} S \subseteq S\}$
- **Note:** By definition, empty set and the whole space are both open and closed
- Bounded set: That is contained in a ball of a finite radius
- Compact set: The one that is closed and bounded
- Couple of useful theorems
  - A compliment of the closed set is an open set, and vice versa
  - A convex combination of interior and closure points is in the interior: $S \subseteq \mathbb{R}^n$, convex with nonempty interior; $x_1 \in \text{cl} S,\ x_2 \in \text{int} S,\ \lambda \in (0,1)$; $\Rightarrow \lambda x_1 + (1-\lambda) x_2 \in \text{int} S$
    - **Corollaries.** If S is a convex set with non-empty interior: closure of S is convex; cl[int(S)] = cl[S] (Closure of the interior of a convex set is just a closure); int[cl(S)] = int[S] (interior of closure is just an interior)
    - **Proof idea**: pick a pt in int(S) near boundary pt (it does exist by definition of the boundary); pick an arbitrary point near the "interim" pt -- close enough (i.e. near the point we are trying to prove to be in the interior); then I can construct a point in the neighborhood of the int point s.t. close enough to be in S and s.t. the point from prev bullet will be a convex combination of the first one => will be in S ! (since it will be a convex combination of two interior points); Hence, this mid-point is an interior point, as has a whole ball around it is in the S!
    - **Sidenote:** examples of convex sets with empty interior: a point in 2D; a hyperplane in 3D; a 3D ball in 5D %-); **side-sidenote** = see what is *relint*

5 Separation results and Farkas' lemma

5.1 Thm. Existence of the min distance between a point and a set
let $S \ne \emptyset, S \subseteq \mathbb{R}^n$ – closed convex set; let $y \in \mathbb{R} \setminus S$; $\Rightarrow \exists!\ \bar x \in S$ with minimum distance between y and S; Moreover, $\bar x$ is such point $\iff \forall x \in S:\ (y - \bar x)^T(x - \bar x) \le 0$

5.2 Thm. Existence of a separating hyperplane (point, set)
let $S \subseteq \mathbb{R}^n$, $S \ne \emptyset$ – closed convex set, $y \in \mathbb{R}^n \setminus S$; $\Rightarrow \exists p \ne o, \alpha \in \mathbb{R}:\ \forall x \in S\ p^T y > \alpha$ and $p^T x \le \alpha$

Corollaries: ∀ closed convex set $S \subseteq \mathbb{R}^n$ is an ∩ of all halfspaces, containing S; Let $S \subset \mathbb{R}^n, S \ne \emptyset, y \notin \text{cl}(\text{conv}(S))$. Then ∃ a hyperplane, separating S and y

5.3 Farkas' lemma & Co.
Let $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{l \times n}$. **Then** exactly **one** of the following systems has a solution:
- Farkas': ① $Ax \le o,\ c^T x > 0$; ② $A^T y = c,\ y \ge o$; Geometric intuition (if c is not in a cone made of rows of A -- there exists a separating hyperplane (which would be x in the first system))

- Gordan's: version 1: ① $Ax < o$; ② $A^T y = o,\ y \ge o,\ y \ne o$; version 2 (essentially, the same): ① $Ax < o$; ② $A^T y = o,\ y \ge o,\ 1^T y = 1$
- Another version: ① $Ax \le o,\ x \ge o,\ c^T x > 0$; ② $A^T y \ge c,\ y \ge o$
- Yet another version: ① $Ax \le o,\ Bx = o,\ c^T x > 0$; ② $A^T y + B^T z = c,\ y \ge o$
- Key facts (the rest are usually proven with these two)

5.4 Supporting hyperplanes and separating a set from a point
- **Def.** A hyperplane, *supporting S at y* (y is assumed to be in the boundary of S): $H = \{x:\ p^T(x - y) = 0\}$ – iff $\forall x \in S\ p^T(x-y) \ge 0$ (or ≤ 0); **Note**: if $S \not\subseteq H$ then H – "proper supporting hyperplane"
- **Thm.** Existence: let $S \subseteq \mathbb{R}^n$, $S \ne \emptyset$ – convex; $\bar x \in \partial S$; $\Rightarrow \exists p \ne o:\ \forall x \in \text{cl}(S)\ p^T(x - \bar x) \le 0$
  - **Corollaries**: $S \ne \emptyset, S \subseteq \mathbb{R}^n$ – convex; $\bar x \notin \text{int}(S) \Rightarrow \exists p \ne o:\ \forall x \in \text{cl}(S)\ p^T(x-\bar x) \le 0$; $S \ne \emptyset, S \subseteq \mathbb{R}^n$, $y \notin \text{int}(\text{conv}(S)) \Rightarrow$ ∃ a hyperplane separating S and y; $S \ne \emptyset, S \subseteq \mathbb{R}^n$, $\bar x \in \partial S \cap \partial\text{conv}(S) \Rightarrow$ ∃ hyperplane separating S from $\bar x$

5.5 Separating sets
- **Definitions**: $H = \{x:\ p^T x = \alpha\}$ – a hyperplane; $S_1, S_2 \subseteq \mathbb{R}^n$
  - **Def.** H *separates* S1 and S2: $\forall x \in S_1:\ p^T x \le \alpha$; $\forall x \in S_2:\ p^T x \ge \alpha$
  - **Def.** H ***properly*** separates S1 and S2: H separates S_1 and S_2; $S_1 \cup S_2 \not\subseteq H$
  - **Def.** H ***strictly*** separates S1 and S2: $\forall x \in S_1:\ p^T x < \alpha$; $\forall x \in S_2:\ p^T x > \alpha$
  - **Def.** H ***strongly*** separates S1 and S2: $\exists \varepsilon > 0$: $\forall x \in S_1:\ p^T x \ge \alpha + \varepsilon$; $\forall x \in S_2:\ p^T x \le \alpha$
  - Strong ⊆ Strict ⊆ Proper ⊆ just "Separation"
- **Thm.** Separation of two non-empty disjoint convex sets: $S_1, S_2 \subseteq \mathbb{R}^n$, $S_i \ne \emptyset$, $S_1 \cap S_2 = \emptyset$, S_i – convex; ⇒ ∃ a separating hyperplane
  - **Cor.** $S_1, S_2 \subseteq \mathbb{R}^n$ – non-empty convex sets; $\text{int}(S_2) \ne \emptyset$, $S_1 \cap \text{int}(S_2) = \emptyset$; ⇒ ∃ a hyperplane, separating S_1 and S_2
  - **Cor.** $S_1, S_2 \subseteq \mathbb{R}^n$, $S_i \ne \emptyset$; $\text{int}(\text{conv}(S_i)) \ne \emptyset$; $\text{int}(\text{conv}(S_1)) \cap \text{int}(\text{conv}(S_2)) = \emptyset$; ⇒ ∃ a hyperplane, separating S_1 and S_2
- **Thm.** Strong separation of closed and convex sets: S_1, S_2 – disjoint closed and convex sets; S_1 – bounded; ⇒ ∃ a hyperplane that strongly separates S_1 and S_2
  - **Cor.** $S_1, S_2 \ne \emptyset$, $S_i \subseteq \mathbb{R}^n$ – bounded; $\text{cl}(\text{conv}(S_1)) \cap \text{cl}(\text{conv}(S_2)) = \emptyset$ ⇒ ∃ a hyperplane that strongly separates S_1 and S_2
- **Note**: $S_1, S_2 \in \mathbb{R}^n$ - convex; ∃ a hyperplane that strongly separates S_1 and S_2 **iff**: $\inf\{\|x_1 - x_2\|:\ x_1 \in S_1,\ x_2 \in S_2\} > 0$

6 Convex functions

6.1 Definitions (+ epigraph)
- $\forall x_1 \ne x_2 \in \text{Dom} f, \lambda \in [0,1]$: $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$
- f is concave iff (-f) is convex
- f over non-empty convex set is convex iff it's *epigraph* is convex: epigraph (everything "above" the graph) $\{(x,y) \in S \times \mathbb{R}:\ y \ge f(x)\}$; hypograph (everything "below" the graph) $\{(x,y) \in S \times \mathbb{R}:\ y \le f(x)\}$
- $f'(\bar x; d) = \lim_{\lambda \to 0+} \frac{f(\bar x + \lambda d) - f(\bar x)}{\lambda}$
- A subgradient: $S \ne \emptyset$, $f: S \to \mathbb{R}$ - convex, $\bar x \in \text{int} S$; $\xi:\ \forall x \in S:\ f(x) \ge f(\bar x) + \xi^T(x - \bar x)$; A set of all subgradients is called "**subdifferential**"; **Thm.** For convex differentiable function subdifferential = a singleton (viz., the gradient): $S \subseteq \mathbb{R}^n$, $S \ne \emptyset$ - convex; $f: S \to \mathbb{R}$ - convex; f - differentiable at $\bar x \in \text{int} S$ $\Rightarrow \partial f(\bar x) = \{\nabla f(\bar x)\}$

6.2 Jensen's inequality
(for any convex f, corollary of the epigraph thm; "generalization" of the def'n?) Assuming lambdas correspond to a convex combination: $f(\sum \lambda_i x_i) \le \sum \lambda_i f(x_i)$; $f(\sum \lambda_i x_i) \le \sum \lambda_i f(x_i)$; $f(\mathbb{E}X) \le \mathbb{E}f(X)$

6.3 Checking convexity
- **6.3.1 By definition (obviously)**
- **6.3.2 Geometric observations**: **Thm.** hyp/epi (or an alternative definition): A function is convex **iff** its epigraph is convex; A function is concave **iff** its hypograph is convex; **Thm.** A function is (s)convex **iff** any linear restriction is (s)convex (i.e. for any x and v: g(t,x,v) = f(x+vt) is convex wrt to t)
- **6.3.3 Gradients** (obviously, for differentiable functions): f is convex **iff** the graph is above its "linear approximation": Dom $f = S$ – convex, open; $\forall x \ne y \in S:\ f(y) \ge f(x) + \nabla^T f(x)(y - x)$; f is convex **iff**: $\forall x_1, x_2 \in S:\ [\nabla f(x_2) - \nabla f(x_1)]^T (x_2 - x_1) \ge 0$
- **6.3.4 Subgradients** (aka FOC): **Subgrad vs. Convexity**: $S \subseteq \mathbb{R}^n, S \ne \emptyset$, convex; $f: S \to \mathbb{R}$
  - A convex function has a subgradient at any interior point of S: f - convex $\Rightarrow \forall \bar x \in \text{int} S\ \exists \xi:\ \forall x \in S:\ f(x) \ge f(\bar x) + \xi^T(x - \bar x)$
  - If a function has a subgradient at any interior pt of S -- it is convex **on the interior of S**: $\forall \bar x \in \text{int} S\ \exists \xi:\ \forall x \in S:\ f(x) \ge f(\bar x) + \xi^T (x - \bar x)$ $\Rightarrow$ f is convex on int S
  - i.e. **for an open S**, a function has a subgradient at any point iff it is convex; If it's closed -- not necessarily! Recall the "Taj Mahal example"
- **6.3.5 SOC[NS]** ("Using Hessians"; **necessary and sufficient**): True if: f – twice differentiable; f – convex **iff**: Dom $f = S$ – convex; $\forall x \in S:\ Hf(x) \succeq 0$ (PSD); A note on *strict* convexity: H is PD --> f is *strictly* convex; f is *strictly* convex =/=> H is PSD (counter-example: x^4)
- **6.3.6 Higher order condition** (**Necessary and sufficient**): $S \ne \emptyset$, open, convex subset of $\mathbb{R}$; f is strictly convex **iff**: $\forall \bar x \in S:\ \exists n$ - even: $f^{(n)}(\bar x) > 0$ and $\forall j \in [2, n-1]:\ f^{(j)}(\bar x) = 0$
- **6.3.7 Useful facts** (i.e. "decomposing" a function into something we know is convex)
  - A **conic combination** of convex functions is convex: f_i - convex, $\alpha_i > 0 \Rightarrow \sum_i \alpha_i f_i(x)$ - convex
  - $f(x, \alpha), \alpha \in \mathcal{A}$ – convex $\to \max_{\alpha \in \mathcal{A}} f(x, \alpha)$ – convex
  - h – convex, g – nondecreasing $\to f(x) = g(h(x))$ – convex
  - g – convex $\to f(x) = g(Ax + b)$ – convex
  - Selected special cases: Any norm is convex (subadditivity + absolute homogeneity); Affine function is both convex and concave (And any convex & concave func is affine)
  - $\forall a, b \ge 0, 0 \le \lambda \le 1:\ a^\lambda b^{1-\lambda} \le \lambda a + (1-\lambda) b$

6.4 Some useful facts
- Any convex function is continuous on the int S: $S \ne \emptyset$, convex; $f: S \to \mathbb{R}$ - convex; ⇒ f is continuous on int S
  - Proof idea: Represent each x as (within certain ball around x-bar): $x = \bar x + \sum_{i=1}^n \alpha_i z_i$, where z's are vectors with the only nonzero element being +1 or -1 in resp. coordinate; Just apply convexity def several times
- Existence of directional derivatives: $f: \mathbb{R}^n \to \mathbb{R}$ - convex, $\bar x \in \mathbb{R}^n, d \in \mathbb{R}^n \setminus o$ ⇒ $\exists f'(\bar x; d)$

6.5 Generalizations of convexity
(Here intuitive description is given followed by a formal definition; see a separate diagram (PDF) on this for details)
- Convexity: "Any line segment between 2 pts on the graph is above the graph"; Any local min over a convex set is actually a global min (Can be proven by contradiction -- assuming there's another opt outside certain ball around x*); $\forall x_1, x_2 \in S, \lambda \in (0,1):\ f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$
- Quasiconvexity
  - Quasiconvex: "A cross section is unimodal or monotonic"; All level sets are convex (An equivalent definition); $\forall x_1, x_2 \in S, \lambda \in (0,1):\ f(\lambda x_1 + (1-\lambda) x_2) \le \max\{f(x_1), f(x_2)\}$
  - Semistrictly quasiconvex ("Strictly quasiconvex" in terms of Bazaraa et al.): "Flatlands are forbidden outside global minimum"; So, any local min is actually a global min (over a convex set; global min is not necessarily unique!); $\forall x_1, x_2 \in S:\ f(x_1) \ne f(x_2), \lambda \in (0,1):\ f(\lambda x_1 + (1-\lambda) x_2) < \max\{f(x_1), f(x_2)\}$
  - Strictly quasiconvex ("Strongly convex" in terms of Bazaraa et al.): "Flatlands are forbidden even for global minimum"; So, obviously, a global min (if exists) is unique; $\forall x_1, x_2 \in S:\ x_1 \ne x_2, \lambda \in (0,1):\ f(\lambda x_1 + (1-\lambda) x_2) < \max\{f(x_1), f(x_2)\}$
- Pseudoconvexity
  - Pseudoconvex: "Gradient points away from min"; So, ∇f = o ⟶ global minimum (although, might be not unique); $\forall x_1, x_2 \in S:\ \nabla^T f(x_1)(x_2 - x_1) \ge 0 \to f(x_2) \ge f(x_1)$; $\forall x_1, x_2 \in S:\ f(x_2) < f(x_1) \to \nabla^T f(x_1)(x_2 - x_1) < 0$
  - Strictly pseudoconvex: Essentially, we forbid a "flatland" in the global min (observe that constant function is not strictly pseudoconvex!); $\forall x_1, x_2 \in S:\ x_1 \ne x_2:\ \nabla^T f(x_1)(x_2 - x_1) \ge 0 \to f(x_2) > f(x_1)$; $\forall x_1, x_2 \in S:\ x_1 \ne x_2:\ f(x_2) \le f(x_1) \to \nabla^T f(x_1)(x_2 - x_1) < 0$

7 Linear methods

7.1 Linear formulation basics
- **7.1.1 "General" form** (not more general than the standard form, actually): $\min c^T x$; (LP) s.t. $x \in P$
- **7.1.2 Standard form**: $\min c^T x$; s.t. $Ax = b,\ x \ge 0,\ x \in \mathbb{R}^n$; $c \in \mathbb{R}^n; A \in \mathbb{R}^{m\times n}; b \in \mathbb{R}^m$
  - Reducing to the st. form: $\max f_0(x) = -\min f_0(x)$; **Inequality** → add slack/surplus variables; **negative** variables → introduce new ones $y = -x$; **free** vars → split into nonnegative $x^+ - x^-$
  - **Geometry:** "lifting" the polytope to a higher dimension (but the dimensionality of the object remains the same!)
- **7.1.3 Canonical form**: Ax=b is in CF **iff** A contains Id-matrix as a submatrix

7.2 Polyhedra, LP+XP/BFS
- **Defs**
  - Basic solution ("BSol"): satisfies all equality constraints; $\text{rank}(A_{I(x)}) \ge n$ — At least *n* of the active constraints at the point are linearly independent
  - Basic feasible solution ("BFS"): Basic solution; Feasible (+satisfies all ineq constraints); **Def.** degenerate BFS -- BFS with > n active constraints
  - Extreme point ("XP"): **Def.** That cannot be represented as a convex combination of other points
  - A note on "representations": Being an XP / B(F)S does depend on representation (e.g. imagine a halfspace -- it does not have any XP, but if we move to standard form -- the origin "becomes" an XP of a new polyhedron, in higher dimensions); If a point is an XP of a polyhedron -- it will stay an XP in higher dimensions; But if it is an XP in higher dimensions -- it can be not an XP in lower dimensions
- **Extreme points and B(F)S** (Property of a point in a polyhedron)
  - **Thm.** x is an XP *iff* it is a BFS: If x is an XP, but not BFS -- we could have found a direction and choose two points on the corresponding line s.t. their convex combination is "XP"; If x is BFS, but not XP -- we can construct a nonzero vector in a null-space of A_I -- that will contradict BFS requirement (on rank(A_I))
  - **Thm.** A nonempty polyhedron has an XP iff it doesn't contain a line: $\exists \text{XP} \Leftrightarrow \nexists x \in P, d \in \mathbb{R}^n:\ \forall \lambda \in \mathbb{R}\ x + \lambda d \in P$; **Note:** a polyhedron in standard form always has an XP -- since it cannot contain a line (it would have "cut through" the x_i=0 hyperplanes)
  - **Thm.** A - full (row) rank, P in std form; P≠∅; then x – BSol **iff:** ("characterization of XP") $Ax = b$ and $\exists m$ lin/indep cols of A – i.e. matrix B of rank m: $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ o \end{bmatrix}$
    - "Basic variables" (for a BSol): **Def.** basic variables -- the ones corresponding to the chosen lin/indep cols (these cols are also called "basic cols". Matrix composed of these -- "basic matrix"); **Def.** nonbasic vars -- vars that are not basic (surprisingly)
  - **Note.** Number of XP (for a polyhedron): is finite - at most $\binom{n}{m}$
  - Degenerate BSol: more than n constraints are active (Like, n=n of 3 lines in 2D or 4 planes in 3D); **Thm.** P is in std for; then x is non-degenerate iff: $x_B > 0$; OR degenerate if more than (n-m) components are zero
- **(Extreme) rays** (XR / XD)
  - **Def.** A *direction of S* (or "ray"): d such that: $d \ne o,\ \forall x_0 \in S,\ \lambda \ge 0:\ x_0 + \lambda d \in S$ (feasible); **Note**: d is a direction of S iff $Ad = 0$ and $d \ge o$
  - **Def.** Extreme direction of S (or Extreme ray): $d = \lambda_1 d_1 + \lambda_2 d_2,\ \lambda_1, \lambda_2 > 0 \to \exists \alpha > 0:\ d_1 = \alpha d_2$; i.e. the one that cannot be represented as a weighted sum of two other distinct directions (up to scaling constants)
  - **Thm.** d is an extreme direction of S (a polyhedron **in standard form**) iff (Characterization of extreme rays aka "algorithmic approach"): A can be decomposed into $[B, N]$ such that $\exists j:\ B^{-1} a_j \le o$ and $\exists \alpha > 0:\ d = \alpha \begin{bmatrix} -B^{-1} a_j \\ e_j \end{bmatrix}$, where e_j – $n-m$ vector of zeros except for a 1 in position j; **Cor.** Number of extreme rays is finite
  - **Thm.** d is an XD of S (a polyhedron in **standard form**) iff: scaled to $1^T d = 1$, it is an XP of a polyhedron: $D = \{d:\ Ad = 0,\ d \ge o,\ 1^T d = 1\}$
- **Representation theorem** (for a point in a polyhedron): Any feasible point of a polyhedron in standard form can be represented as a convex combination of XPs plus conic combination of XRs: $x = \sum_{j=1}^k \lambda_j x_j + \sum_{i=1}^l \mu_i d_i$; $\sum \lambda_j = 1,\ \lambda_j \ge 0,\ \mu_i \ge 0$; $x_1, \ldots, x_k$ – are all XPs of the polyhedron; $d_1, \ldots, d_l$ – are all XRs of the polyhedron; Sometimes people speak about the **V**-representation of a polyhedron (with **V**ertices and directions) vs. H-representation (with **H**alfspaces - Ax <= b)
- **An LP (min) can be, in principle:** infeasible (p* = inf); unbounded (p* = -inf); solvable, p*> -inf; if there is an XP -- there is an optimal XP (BFS)
  - **Thm.** If P – polyhedron w/at least 1 XP either (LP) is unbounded, or ∃ an XP that is optimal: Starting from arbitrary point, if it is not a BFS -- we can choose a direction of cost-non-increase and keeping feasibility -- and travel until we "meet" another constraint (and increase rank(A_I)) (So we either find "infinitely good" direction, or bump into an optimal BFS)

7.3 Simplex method (includes sensitivity)

7.3.1 Geometry (simplex in a nutshell)
- Pick any XP (BFS) of the polyhedron: Solve an auxiliary LP for that (that is easier to solve); if the problem is infeasible -- we'll see it here
- Move to an adjacent XP (BFS) -- if it improves objective
- If no such move possible -- it is an optimal solution
- If we can take off along some "edge" into infinity, improving the objective -- the problem is unbounded
- Repeat

7.3.2 The Tableau
- **Cols** -- coordinates of A_j in the current basis (so, if you see a col from an identity matrix -- this col corresponds to a basic variable)

$-c_B^T B^{-1} b$	$\bar c = c^T - c_B^T B^{-1} A$
$B^{-1} b$	$B^{-1} A$

- Top left corner -- objective value with the negative sign; "Top right" row -- reduced costs; Bottom right block -- coefficients in the current basis, B; Bottom left block -- current values for basic variables
- Pivot: $x'_B = \begin{bmatrix} \beta_1 \\ \ldots \\ \beta_m \end{bmatrix} + x_k \begin{bmatrix} -\alpha_1 \\ \ldots \\ -\alpha_m \end{bmatrix}$ ($B^{-1}b$, $B^{-1}A_k$); The first term are old values of basic vars; x_k -- is the entering variable; Col in the second term -- corresponding col of the coefficients block (corresponding to the entering variable) ... here!; **Note:** min ratio test just ensures that all x's remain nonnegative (i.e. feasibility)
- Optimality criterion: $\bar c \ge 0$ (otherwise I can improve); $B^{-1} b \ge 0$ (feasibility); **Thm.** Consider BFS x, associated with basis B and reduced costs c-bar: if $\bar c \ge 0$ – x is optimal; if x is optimal and non-degenerate: $\bar c \ge 0$
- **Nice things we can read from the tableau** (More or less directly):
  - "RHS" col -- values of basic variables in current BFS
  - "corner" value -- objective at current BFS (note: depending on the notation, the sign might be different -- be careful!)
  - If you know the initial BFS, and the coefficients matrix included identity matrix as a submatrix (e.g. when we have slack variables) -- B^(-1) can be recovered
  - top row -- reduced costs
  - non-positive col corresponding to a negative reduced cost -- unboundedness direction (beware: signs!): take these values, corresponding to current basic vars, with inverted sign; put 1 for var, corresponding to this column (i.e. to this non-basic variable); put zeroes for all non-basic variables; Package and sell as an R^n vector (... from ... this is easy to see...)
  - top row after some magic -- optimal dual variables: If you had an identity sub-matrix somewhere within A in first place (say, if we have m slack variables); Note that $y^{*T} = c_B^T B^{-1}$; So, if we subtract a row of costs from the top row – we end up with numbers from $-c_B^T B^{-1}$ somewhere among these; In particular, if we had slack variables -- their reduced costs in optimal tableau are actually (sign-inverted) optimal dual variables corresponding to these constraints (since (if) costs were zero)

7.3.3 Possible outcomes
- **Infeasibility**: Is usually seen from the first-stage problem; If all the basic solutions are infeasible -- the whole problem is infeasible (if it helps -- we could just check all n-choose-m points); Note, that a simple system feasibility sometimes is easier to check with Farkas' lemma (has nothing to do with simplex method, admittedly)
- **Unboundedness**: At any point, if the reduced cost is negative -- I can improve by moving along this direction (i.e. by doing a pivot with this variable); How far can I go? Well, until I "meet" the closest non-negativity constraint (I don't care about the rest, since I am moving along directions that ensure Ax=b); Can it happen that I never meet it -- so the problem is unbounded? **Yes**: if all alphas are nonpositive, obviously; i.e. **Criterion**: negative reduced cost, nonpositive column
- **Optimality**: When we have just arrived at an optimal tableau; If we cannot move anywhere from this point without increasing the objective value -- this is the only optimal solution
- **Alternative optimal solutions**: Start with an optimal tableau; If we have zero-reduced-cost non-basic variables -- there is a direction, that potentially do not damage our objective; On top of it, if I technically pivot (change the basis) -- then there is an *alternative basis*, corresponding to an *optimal point*.; On top of it, in turn, if I can actually move along this direction (min ratio test does not prevent me from) -- there is an alternative (optimal) BFS (and hence, a whole lot of (infinitely many) alternative optimal solutions -- since any convex comb of opt sol is an opt sol)

7.3.4 Degeneracy
- In degenerate cases (at least 1 basic variable is zero) -- a pivot does not necessarily lead to a different point
- Moreover, sometimes this might result in an infinite loop: The simplest example that I'm aware of includes 7 variables; Also see notes on convergence below (on what to do)
- **Note:** degenerate BFS does not necessarily mean a redundant constraint! (the constraint might be redundant just for this particular point)

7.3.5 Pivoting rule
- How do we choose entering and leaving variables, if we have options? (say, several nonbasic variables have negative reduced costs and we can technically pivot there)
- **Dantzig rule:** choose the "most negative" reduced cost-variable to enter; break ties arbitrarily (kind of "greedy" rule)
- Some alternatives: **Bland's rule:** choose an eligible variable with the smallest subscript (guarantees to avoid cycling); **Random rule:** choose an eligible variable randomly

7.3.6 Finding initial BFS
(To run a simplex -- we need an initial BFS, which is not always obvious)
- Two phase method: $\min 1^T y$ s.t. $Ax + y = b,\ x, y \ge o$; Opt value >0 => initial P is infeasible; Opt value = 0 => found a BFS: If y are in the basis -- pivot them out; If y are not in the basis -- disregard these cols and take the BFS to the initial problem; Then just solve the initial problem starting from the found BFS
- Big-M method (M is a "very big number" (larger than anything finite, by definition)): $\min c^T x + M 1^T y$ s.t. $Ax + y = b,\ x, y \ge o$
- **Note:** these two seem to take exactly the same steps; but computationally, 2-phase is usually better (less prone to comp errors, etc.) (i.e. same steps -- under the Dantzig rule)

7.3.7 Sensitivity analysis
(essentially, based on the tableau; Based on Bertsimas book)
- Basic conditions: $B^{-1} b \ge 0$; $\bar c = c^T - c_B^T B^{-1} A \ge 0$
- Variables
  - Adding a new variable: Same BFS is optimal = reduced cost for the new var is nonnegative: $c_{n+1} - c_B^T B^{-1} A_{n+1} \ge 0$
  - Deleting a nonbasic variable: nothing changes, essentially (just adjust dimensionalities, as appropriate)
- Changing RHS: Objective change: $obj = c_B B^{-1} b$; $\Delta(obj) = c_B B^{-1} \delta = y^{*T} \delta$; BFS: $x_B = B^{-1} b \ge 0$; $x_B' = x_B + B^{-1} \delta b$
- Changing costs: Primal feasibility is unchanged, obviously; Nonbasic cost changed (conditions for the basis to remain optimal): $c_B^T B^{-1} A_j \le c_j + \delta$; In fact, change in costs $\delta \ge -\bar c_j$; Basic cost changed (again, condition for the basis to remain optimal): $(c_B + \delta e_i)^T B^{-1} A_j \le c_j\ \ \forall j \ne i$; q_l being the lth entry of $B^{-1} A_j$: $\delta q_l \le \bar c_j\ \ \forall j \ne i$
- Constraints — Adding a new constraint
  - Equality: Old optimal BFS does not violate it -- no changes needed (opt dual for the new one = 0); Old optimal BFS violates this one -- we'll need to solve an aux LP to find a new solution (see Bertsimas for details; essentially, it's like adding a big M to objective for the slack/surplus of the new constraint -- and solving the problem by optimality)
  - Inequality: If current sol is feasible -- nothing changes (corresponding new opt dual is zero); If not: expand the basis by one more var (say, a slack for a new constraint); pivot to get to the same basis (Note it's just BSol, not BFS); Get feasibility back -- dual simplex update
- Changes in A (coefficients): Nonbasic cols: just think about deleting (nothing changes) and then adding a new nonbasic variable (see above); *Basic cols -- more complicated case; both feasibility and optimality are affected*
- The story with "allowable ranges": Observe that some expressions for objective, BFS values, etc. above are calculated assuming the **basis B remained the same**; This is a common assumption, but it implies that this analysis is applicable only within certain ranges of changes, that keep B unchanged -- see formulae for that above; Usually, whatever the changes at hand are -- the ranges should be easy(tm) to compute from: BFS feasibility – $B^{-1} b \ge o$; nonnegative reduced costs for all nonbasic vars; (depending on what is affected)

7.3.8 On convergence
- With a Dantzig rule, it was shown that the algorithm can visit all the vertices of certain polytope -- so, exponential (the keyword is "Klee-Minty cube")
- I am not aware if it was shown that there exists a pivoting rule that makes the process polynomial
- There is some research on expected runtime of simplex; in general -- appears to be good in practice (Borgwardt (1982) -- exp. runtime is polynomial, under certain assumptions)
- Without proper care, the algorithm can stuck in a degenerate point. To mitigate: Make a small perturbation of the right-hand side (b) (note that this makes the problem relatively rare in practice); Change pivoting rule (see sample alternatives above) (e.g. when we suspect or know that we are stuck)

7.3.9 On implementations
- Naive — Simplified algo (typical iteration) (computationally heavy operations are marked with !): Form a basis B; ⚠ compute $p^T = c_B^T B^{-1}$; Calculate reduced costs: $\bar c^T = c^T - p^T A$ (Select a col to enter the basis, as usual); ⚠ find $u = B^{-1} A_j$ — To determine the leaving variable; Determine which var exits the basis; Find new BFS; and so on; $O(m^3 + mn)$
- Revised — **Idea:** just keep inverted B separately -- essentially, this is kind of a full tableau flavor, but we keep B-inverse separately (update it cheaply -- and then just multiply matrices by vectors); Imagine that we just keep B-inverse and don't keep A times B-inv -- calculate its parts as necessary (and can use less space and, for some iterations, save some time -- if we don't need to compute all the reduced costs always to pivot (depending on the pivot rule)); Also, if we pivot in, say first negative-reduced-cost var -- some iterations can be faster than in full-tableau
  - Simplified algo (typical iteration): Start with B and B^{-1}; compute $p^T = c_B^T B^{-1}$; Calculate reduced costs: $\bar c^T = c^T - p^T A$ (Select a col to enter the basis, as usual (say, #j)); find $u = B^{-1} A_j$ — To determine the leaving variable: if no component is positive -- the problem is unbounded; if there is a positive component -- run a min ratio test: $\theta^* = \min_{\{i: u_i > 0\}} \frac{x_{B(i)}}{u_i}$; Update basic vars values: $y_j = \theta^*$; $y_i = x_{B(i)} - \theta^* u_i$; update B-inv: Form a matrix $[B^{-1}|u]$; do elementary row operations to make u a unit vector e_l; The first m cols of that matrix is the new B-inverse
  - $O(m^2 + mn)$
- Full-tableau — **Idea:** essentially, same logic -- but don't keep B-inverse separately (see the Tableau section for details); $O(m^2 + mn)$

7.4 Existence of a finite solution vs. "unboundedness" directions
(P) has a finite solution (for a non-empty polyhedron) iff there are no "unboundedness directions". And if a finite solution does exist -- the optimal value is achieved in an XP, i.e.: $\min c^T x$; (P): S.t. $Ax = b,\ x \ge o$; Finite opt exists iff $\forall d_i$ – XR: $c^T d_i \ge 0$

7.5 Complementary slackness conditions (repeated)
(the way we had it in lectures) $A^T y \le c \leftrightarrow A^T y + s = c$; $s_i^* x_i^* = 0$; **Note:** if we have feasible solutions to P and D -- then they are optimal iff corresponding complementary slackness conditions hold (i.e. both, for primal and dual)

7.6 Duality in linear case
- Primal-dual conversion — Simple case: **Primal:** $\min_{x \in \mathbb{R}^n} c^T x$ s.t. $Ax = b,\ A \in \mathbb{R}^{m\times n}$, $x \ge 0$; **Dual:** $\max_{y \in \mathbb{R}^m} b^T y$ s.t. $A^T y \le c$, y – free
- General case:

PRIMAL	minimize	maximize	DUAL
constraints	$\ge b_i$	≥ 0	variables
	$\le b_i$	≤ 0	
	$= b_i$	free	
variables	≥ 0(sensible)	$\le c_j$(sensible)	constraints
	≤ 0(bizarre)	$\ge c_j$(bizarre)	
	free(odd)	$= c_j$(odd)	

- Note the possible instance types for P vs D:

		Primal		
		Finite opt.	Unbounded	Infeasible
Dual	Finite opt.	✓	x	x
	Unbounded	x	x	✓
	Infeasible	x	✓	✓

- **Weak duality** : $\forall y \in Y, x \in X:\ \ c^T x \ge b^T y$
- **Strong duality** : $d^* = p^*$

8 Nonlinear methods

8.1 Key (geometric) definitions (for differentiable functions, obviously)
- **8.1.1 Objective descent**: $F(\bar x) = \{d:\ \exists \delta > 0:\ \forall \lambda \in (0, \delta)\ f(\bar x + \lambda d) < f(\bar x)\}$; "Approximations": $F_0(\bar x) = \{d:\ \nabla f(\bar x) d < 0\}$; $F'_0(\bar x) = \{d \ne o:\ \nabla f(\bar x) d \le 0\}$; $F_0 \subseteq F \subseteq F'_0$; Refinements: f – pseudoconvex at $\bar x \to F_0 = F$; f – strictly pseudoconvex at $\bar x \to F = F'_0$
- **8.1.2 Feasible directions**: $D(\bar x) = \{d:\ \exists \delta > 0:\ \forall \lambda \in (0,\delta):\ \bar x + \lambda d \in S\}$; "Approximations": $G_0(\bar x) = \{d:\ \forall i \in I\ \nabla g_i(\bar x) d < 0\}$; $G'_0(\bar x) = \{d \ne o:\ \forall i \in I\ \nabla g_i(\bar x) d \le 0\}$; $G_0 \subseteq D \subseteq G'_0$; Refinements: $g_i, i \in I$ – strictly pseudoconvex at $\bar x \to D = G_0$; $g_i, i \in I$ – pseudoconcave at $\bar x \to D = G'_0$

8.2 Key expressions
- **Fritz-John (FJ) Necessary condition**: $\bar x$ is feasible; $\exists u_0, u_i, i \in I, v_j, j = \overline{1,l}$: $u_0 \nabla f(\bar x) + \sum_{i \in I} u_i \nabla g_i(\bar x) + \sum_{i=1}^l v_i \nabla h_i(\bar x) = 0$; $(u_0, \bar u, \bar v) \ne o$, $(u_0, \bar u) \ge o$
  - **Geometry**: "Farkas counterpart" of: $F_0 \cap G_0 = \emptyset$; i.e. (-∇f) is not in the cone of ∇ (binding constraints) -- or zero (we cannot move along straight line s.t. improve objective and remain feasible)
  - A valid **necessary** condition for local optimality (but sometimes is too broad) (e.g. we always can add a redundant constraint to make any feasible point FJ)
- **Karush-Kuhn-Tucker (KKT) condition**: Primal feasibility **(PF)** $g_i(x^*) \le 0\ \ i = \overline{1,m}$ and $h_i(x^*) = 0\ \ i = \overline{1,p}$; Dual feasibility **(DF)**: $\lambda_i \ge 0\ \ \forall i = \overline{1,m}$; $\nabla_x L(x^*; \lambda, \mu) = \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{i=1}^p \mu_i \nabla h_i(x^*) = 0$; Complementary slackness **(CS)** $\lambda_i f_i(x^*) = 0\ \ i = \overline{1,m}$; Actually, this is FJ for u0>0 case (i.e. we exclude the scenario "∇f = 0 => condition holds"); $\bar x$ – KKT point $\leftrightarrow F_0 \cap G' = \emptyset$

8.3 Necessary conditions
(i.e. if x-bar is a local minimum, **THEN:**)

Type	Constrained	Unconstrained
FOC	$F_0 \cap D = \emptyset$, i.e. $\forall d \in D:\ \nabla^T f(\bar x) d \ge 0$; $F_0 \cap G_0 \cap H_0 = \emptyset$; $\bar x$ – FJ point (also with eqty constraints); $\leftrightarrow F_0 \cap G_0 = \emptyset$; recall: $G_0 \subseteq D$; if $\nabla g_i, i \in I, \nabla h_i, i = \overline{1,l}$ – linearly independent, h_i – continuously differentiable → $\bar x$ – KKT point	$\nabla f(\bar x) = o$ (since then $F_0 = \emptyset$; works for open S as well)
SOC	–	$\nabla f(\bar x) = o$, $Hf(\bar x)$ – PSD
Higher order	–	**Multivariate:** f – infinitely differentiable, For $f(\bar x + \lambda d)$: 1st nonzero derivative term if exists, should correspond to even power of λ

Notes:
- $\exists \varepsilon > 0:\ N_\varepsilon(\bar x) \cap S$ has empty interior → $\bar x$ if optimal – is necessarily FJ
- Even for convex programs -- KKT is not necessary, generally speaking
- **Rule of thumb:** CQ make KKT necessary
- **Note:** a nice counter-example for the "sufficiency" of the nonzero-derivatives criterion -- Peano function (two "embedded" parabolas)

8.4 Sufficient conditions
(i.e. if this holds -- then x-bar is an optimum)

Type	Constrained	Unconstrained
FOC	**local min :** f – convex, → $\bar x$ – local min **iff** $\exists \xi$ - subgradient of f : $\forall x \in S:\ \xi^T(x - \bar x) \ge 0$ **local min :** • $F_0 \cap G_0 = \emptyset$ (no eqty constraints) • f – pseudoconvex at $\bar x$, • $\exists \varepsilon > 0:\ \forall x \in N_\varepsilon(\bar x) \cap S:\ (x - \bar x) \in D$ **local min :** • $F_0 \cap G_0 \cap H_0 = \emptyset$, • f – pseudoconvex at $\bar x$, • $g_i, i \in I$ – strictly pseudoconvex over $N_\varepsilon(\bar x)$, • h_i – affine **local min :** • $\bar x$ – FJ point, S – relaxed* feasible region, • $\exists \varepsilon > 0$: f - pseudoconvex over $N_\varepsilon(\bar x) \cap S$ at $\bar x$; • $g_i, i \in I$ - strictly pseudoconvex over $N_\varepsilon(\bar x) \cap S$; • $h_i, i = \overline{1,l}$ – affine, ∇h_i – linearly independent **local min :** • $\bar x$ – KKT point, • h_i – quasiconvex for $i \in J^+$ and quasiconcave for $i \in J^-$, • $\exists \varepsilon > 0$ such that : ▷ f – pseudoconvex over $N_\varepsilon(\bar x) \cap S$ ▷ $g_i, i \in I$ – differentiable and quasiconvex over $N_\varepsilon(\bar x) \cap S$ **global min :** • $\bar x$ – FJ point, S – relaxed* feasible region, • f - pseudoconvex at $\bar x$; • $g_i, i \in I$ - both strictly pseudoconvex and quasiconvex at $\bar x$; • $h_i, i = \overline{1,l}$ – affine, ∇h_i – linearly independent **global min :** • $\bar x$ – KKT point, • h_i – quasiconvex for $i \in J^+$ and quasiconcave for $i \in J^-$, • f – pseudoconvex at $\bar x$ • $g_i, i \in I$ – differentiable and quasiconvex at $\bar x$	**global min :** f – (pseudo)convex, $\bar x$ – min **iff** ∃ zero (sub)gradient
SOC	**local min :** $\nabla^2 L$ is PSD $\forall x \in N_\varepsilon(\bar x) \cap S$ **strict local min :** $\bar x$ – KKT point, $\nabla^2 L(\bar x)$ is PD or $\forall d \in C:\ d^T \nabla^2 L(\bar x) d > 0$ **global min :** $\nabla^2 L$ is PSD $\forall x \in S$	**strict local min :** $\nabla f(\bar x) = o$; $Hf(\bar x)$ – PD
Higher order	see below	

where:
- $I = \{i:\ g_i(\bar x) = 0\}$, $I^+ = \{i \in I:\ \bar u_i > 0\}$, $I^0 = \{i \in I:\ \bar u_i = 0\}$
- $J^+ = \{j:\ \bar v_j > 0\}$, $J^- = \{j:\ \bar v_j < 0\}$
- $\bar x$ - a KKT point with corresponding multipliers $\bar u_i$ and $\bar v_i$
- $\nabla^2 L(\bar x) = \nabla^2 f(\bar x) + \sum_{i \in I} \bar u_i \nabla^2 g_i(\bar x) + \sum_{i=1}^l \bar v_i \nabla^2 h_i(\bar x)$
- $C = \{d \ne o:\ \nabla^T g_i(\bar x) d = 0, i \in I^+;\ \nabla^T g_i(\bar x) d \le 0, i \in I^0;\ \nabla^T h_i(\bar x) d = 0, i = \overline{1,l}\}$
- Obviously, f,g and h are assumed twice differentiable, where necessary
- Relaxed feasible region S* = with dropped non-binding constraints

For infinitely-differentiable univariate function (**necessary and sufficient** condition for local optimality): $\bar x$ - local min ↔ first non-zero derivative, if exists, is positive and corresponds to an even power of x, i.e.: either ① $f^{(j)}(\bar x) = 0,\ \forall j = 1, \ldots$; or ② $\exists n \ge 2$ – even : $f^{(j)}(\bar x) = 0\ \forall j \le n-1$ and $f^{(n)}(\bar x) > 0$

Note: any convex differentiable function is pseudoconvex

Rule of thumb: what makes KKT sufficient (all of these at the same time; maybe over certain region): ***objective (pseudo)convexity***; Constraint qualifications (so, KKT are necessary); ***constraints "kind-of-convexity"*** (like, quasiconvexity for inequality constraints; quasiconvex+quasiconcave for h's (equalities), etc. -- some details are above)

8.5 A set of alternative optimal solutions (characterization)
$\min f(x)$ s.t. $x \in S$; f - convex, twice differentiable; S - convex; $\bar x$ – an optimum; $\Rightarrow S^* = \{x \in S:\ \nabla^T f(\bar x)(x - \bar x) \le 0,\ \nabla f(x) = \nabla f(\bar x)\}$

8.6 Specialization of KKT to LPs
(Based on Bazaraa et al., Example in the sec. 4.2,14)
- Consider an LP: $\min c^T x$ s.t. $Ax = b,\ Ix \ge o$
- Its' dual is: $\max b^T y$ s.t. $A^T y \le c$, y - free
- Note that KKT conditions for this are: $Ax = b,\ x \ge o$; $A^T y + v = c,\ v \ge o,\ y$ - free; Indeed, correspond to the dual problem feasibility!; $\forall i = \overline{1,n}:\ v_i x_i = 0$
- KKT are necessary and sufficient (if the LP is feasible): $\bar x$ is a KKT point with Lagrange multipliers $(\bar y, \bar v) \Leftrightarrow \bar x$ and $\bar y$ are optimal to the primal and dual, respectively

8.7 Practical considerations -- general "algorithm"
1. Show that (P) has an optimal solution (Say, using the Weierstrass thm)
2. Check conditions so that KKT is necessary (Say, SD (Convexity+Slater's?) + objective convexity)
3. Solve the KKT system -- obtain KKT points (as candidates for optimal solution)
4. Compare objective value for KKT points (Choose the best one)

Rules of thumb
- **Objective** is convex: S.O. : maybe; KKT : not necessary
- **Constraint qualifications** hold: S.O. : maybe; KKT : necessary
- Both: S.O. : always; KKT : necessary sufficient

9 Duality theory
(A sketch on...)
- General idea: **Building bounds** -- create another problem ("dual") s.t. any feasible point gives a bound and (hopefully) optimal value is the same ("strong duality", if holds)
- Lagrangian dual: **Lagrangian :** $L(x; \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$; **Lagrangian dual func :** $\theta(\lambda, \mu) = \inf_{x \in D} L(x; \lambda, \mu)$ (**Note:** always concave. It's just a pointwise inf of a family of affine functions); **Primal problem :** $p^* = \inf_{x \in D} \sup_{\lambda \ge 0, \mu} L(x; \lambda, \mu)$; **Dual problem :** $d^* = \sup_{\lambda \ge 0, \mu} \theta(\lambda, \mu) = \sup_{\lambda \ge 0, \mu} \inf_{x \in D} L(x; \lambda, \mu)$
- Weak duality : $d^* \le p^*$
- Strong duality (SD): ! Definition $d^* = p^*$
  - 2 Sufficient SD Conditions: **Convexity + CQ**
    - CQ = **Slater's condition**: $\exists x \in \textbf{relint} X$, i.e. $x \in X$ s.t. $\forall j = \overline{1,p}\ \ h_j(x) = 0$ (always); $f_i(x) \le 0, i = \overline{1,k}$ and $f_i(x) < 0, i = \overline{k+1,n}$ (if first k inequalities are specified with affine $f_i(x)$'s); **Note** that Slater's condition in case of affine constraints reduces to feasibility!
    - CQ = **LCQ** (Linear CQ): Problem is feasible (not necessarily strictly); all constraints are affine
    - CQ = **LICQ** (Linear Independence CQ): $\nabla g_i(x^*)$ for $i \in \mathcal{A}$ and $\nabla h_i(x^*)$ for $i = \overline{1,p}$ are **linearly independent**
  - 3 Notes: Holds for any LP (until both primal and dual are infeasible) (by the weaker Slater's condition -- applied to primal and dual problems); SD holds for any convex function over a polyhedron
- Linear case (for more details -- see the "Linear methods" part): **Primal:** $\min_{x \in \mathbb{R}^n} c^T x$ s.t. $Ax = b,\ A \in \mathbb{R}^{m\times n}$, $x \ge 0$; **Dual:** $\max_{\mu \in \mathbb{R}^m} -b^T \mu$ s.t. $A^T \mu + c \ge 0$, μ – free; **<- Alternatively ->** $\max_{\nu \in \mathbb{R}^m} b^T \nu$ s.t. $A^T \nu \le c$, ν – free
- Notes: The dual problem is always convex!

10 Literature, resources

- Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 2006. Nonlinear programming: theory and algorithms, [e.g. 3rd ed. ed. Wiley-Interscience, Hoboken, N.J.]. (nonlinear part)
- Boyd, S.P., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press. (convex part)
- Bertsimas, D., Tsitsiklis, J.N., 1997. Introduction to linear optimization, Athena Scientific series in optimization and neural computation. Athena Scientific, Belmont, Mass. (linear part)

Selected Clemson University courses (on these topics):
- **IE8800** - Nonlinear (in 2019 was taught with the Bazaraa book)
- **IE8030** - Linear (an introduction)
- **MATH 8100** -- linear + intro to nonlinear (mostly convex) (in 2018 was taught with Bertsimas + Boyd books)
- **MATH 8110** -- nonlinear