Optimisation Exam Tips

General advice

Be careful about which conditions are sufficient, which are necessary

- there can be many versions of a similar theorem

Be careful whether the optimisation problem is maximisation or minimisation

Unconstricted Optimisation

Maximising f is equivalent to minimising -f

- So usually considering minimisation gives conditions for maximisation for free.

you can optimise multiple variables one by one

- if x, y are free variables, fix x, find optimal y in terms of x
- substitute that into the target function and now minimise for x

It is possible that differentiation CANNOT give all global/local minima/maxima due to differentiability problems.

Induced norms defined in this course is slight different:

Induced Norms. Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the induced matrix norm $\|\mathbf{A}\|_{a,b}$ (called (a, b)-norm) is defined by

$$\|\mathbf{A}\|_{a,b} = \max_{\mathbf{x}} \{\|\mathbf{A}\mathbf{x}\|_b : \|\mathbf{x}\|_a \leq 1\} ,$$

from where it follows that

$$\|\mathbf{A}\mathbf{x}\|_b \leq \|\mathbf{A}\|_{a,b}\|\mathbf{x}\|_a .$$

- inequality $||x|| \leq 1$ is used, instead of $||x|| = 1$. Though they are equivalent, pls use the definition from the book!!

Notes on Definiteness

everything about definiteness is built up **SYMMETRIC** matrices

- if A is not symmetric, $(A + A^T)/ 2$ will be

Properties of positive definite matrices (also applies to negative definite matrices)

- If a matrix is positive definite, so are all the symmetric sub-matrices

- positive definite matrix must be invertible
  - and inverse of positive definite matrix is positive definite
- A, B positive (semi-)definite ⇒ A+ B is positive (semi-)definite
  - so we can break down complicated matrices and judge definiteness of each component
- diagonal entries of positive definite matrix are all positive
- element with largest absolute value must be on the main diagonal
- det(A) > 0

from definition it is trivial that if A is positive (semi-)definite, then -A is negative (semi-)definite

Symmetric matrix A is positive definite ⇔ exists invertible M s.t. $A = MM^T$
A^TA must be positive semi-definite for any matrix A
- if A is full rank, A^TA is positive definite

Criteria
for 2 x 2, directly use tr(A), det(A)
- sum of E.values = tr(A), product of e.values = det(A)

diagonal dominance:
useful for any symmetric matrix, easy to check
but it is **NOT NECESSARY** condition for ddefiniteness
- you must ensure all elements are all positive/non-negative for positive definiteness/semi-definiteness
- A matrix with only weak diagonal dominance may also be definite

principal minor: CANNOT be used to conclude semi-definiteness. But principal minors strictly greater than 0 is equivalent to positive definiteness

if above two criterion fail, try some vectors to estimate the property of the matrix
if not definite

Eigenvalues criteria for indefinite matrix are useful
- but sometimes find two x s.t. $x^TAx > 0$ and $x^TAx < 0$ is easier to show indefiniteness

Optimality conditions

DONT forget to verify SECOND-order conditions !!!

All second-order conditions (including criteria for saddle points) requires two assumptions:

- f is twice continuously differentiable on the domain
- x* is stationary.

these assumptions serve for linear approximation theorem.

Coerciveness to confirms existence of global minima. (or -∞ gives global maxima)

- as long as domain is **closed**, non-empty.
- so if there is only one stationary point, it must be global minima.

Extrema behaviour of quadratic function is basically determined by matrix A

- coerciveness is also determined by A
- b just determines the location of stationary point x
- c only affects the extrema function value

quadratic functions may not come in usual form in questions, it is useful turning them to matrix-vector form.

Indefinite Hessian ensures saddle point, but saddle points appear in other case, i.e. semi-definite Hessian cases

- find two trajectories where the point is maxima on one trajectory, and minima on the other
- (t^a, t^b) may be suitable for x* = (0, 0)
- but if all trajectories give maxima/minima, this point is a local maxima/minima

Least square

||x|| is NOT differentiable at 0, but $||x||^2$ is always differentiable. So square is used for optimal solution to Ax=b instead of ||Ax-b||.

λ: non-negative, larger λ, higher penalty. (but may lead to overfitting)

Gradient descent

zig-zag behaviour ONLY applies to the step size strategy: **exact line search**.

To prove a constant L is the best(smallest) Lipschitz constant of f, find a pair (x, y) s.t.
||f(x)-f(y)|| = L||x - y||

- you may start by trying y = 0

Convexity
Characterisation of convex functions by positive semi-definiteness of Hessian ONLY applies to:
- TWICE continuously differentiable functions
- defined on **OPEN** convex set.

multiplying by POSITIVE constant preserves convexity, not negative.

point-wise maximum of convex function is convex, point-wise minimisation of concave function is concave.

partial minimisation preserves convexity if:
- it is minimised over a convex set

Level sets of convex functions $\{f(x) \leq a\}$ are convex, but $\{f(x) = a\}$ is NOT convex
- However, if f is affine, then $\{f(x) = a\}$ is convex (as it is hyper-plane)

Proving a function is convex
Strategy 1: Try to break down the function, and recognise possible basic convex functions we learnt
Strategy 2: find Hessian and prove it is positive semi-definite
Strategy 3: Write down Gradient inequality, or the monotonicity inequality and try to prove them (works especially well for lower dimensions)

Basic convex functions:
- Norm
- Affine functions $f(x) = a^T x + b$ (affine function is also concave)
- Quadratic with positive semi-definite A

Combination rules:
- any linear combination with positive coefficients
- linear change $g(y) = f(Ay+b)$
- composition of g, f where f, g convex, g is non-decreasing on Im(f)
- pointwise maximisation/ partial maximisation
- partial minimisation over convex set

Convex Optimisation
for orthogonal projection to C, $P_C(x)$, if $x \in C$, $P_C(x) = x$!!!!

when you are dealing with vectors, $\mathbf{x} \geq \mathbf{0}$, $\mathbf{x} \neq \mathbf{0}$ NOT imply $\mathbf{x} > \mathbf{0}$

KKT

After finding Lagrange multipliers, remember to verify λ_i are **positive**

- this is only required for multipliers on **inequality** constraints

Except the Lagrangian = 0 and slater condition, also remember to check **non-negativity** of the Lagrange multipliers associated with inequality constraints.

you can scale any multiplier in KKT by POSITIVE constant to make calculations easier

For LCP (equality constraints allowed):
necessity of KKT only requires f to be continuously differentiable

- but for any vector satisfying KKT to be the optimal solution, you need f to be convex.
- Otherwise, just find the function values of the multiple vectors satisfying KKT and pick the **x** giving smallest f(**x**) (remember to check **x** satisfies all constraints)
  - remember to use Weierstrass theorem to justify existence of global optimal solution.

Duality

Different duality problem construction yields different duality gap! (gap = f* - q*). You can try to pick underlying CONVEX set X to be values satisfying some of the constraint(s). The aims are:

- make Slater's condition satisfied
- make f, g_i convex over X

so that f* = q*.