CS 754 2022 Midsem

Q1 Only instructions

Q2 κ_s is computed by enumerating only 2000 subsets as enumerating all of them is not possible.

The λ_{min} computed over these 2000 subsets will be $\geq$ λ_{min} value computed over the complete set of subsets.

The λ_{max} value computed over these 2000 subsets will be $\leq$ the λ_{max} value computed over the complete set of subsets.

As $\kappa_s = \max_{\beta} \max(1-\lambda_{min}, \lambda_{max}-1)$ it must be less than or equal to the true RIC, ie δ_s.

The correct answer is option # 1:

$\kappa_s \leq \delta_s$

Q3 The global rotation/refl. ambiguity means that the projection angles in this problem [or the corresponding rotation/reflection matrices] can be computed only upto a global unknown rotation/reflection matrix.

That is, if the true rotation matrices are $R_1, R_2, \ldots, R_N$, then any algorithm will give us only

$QR_1, QR_2, \ldots, QR_N$ where Q is

some arbitrary rotation/reflection matrix.

The ambiguity arises because there is no way to decide which of the N macromolecules/particles are in "canonical" position. The tomographic projection of a 3D structure in angle θ = tomographic projection of the same structure rotated by ϕ but taken in angle $\theta + \phi$. See slide 72 of the slides on tomography.

Q4 The projection $g(\rho, \theta)$ will also be bandlimited due to the Fourier slice theorem because its Fourier transform =

a central slice through the Fourier transform of f. The FT values of f beyond any frequency with $|u|$ or $|v| \geq b$ will be 0 in value. This will cause the F.T. of g to also be zero beyond some frequency index.

Q5 For CASSI, we have the foll. forward model for single frame

$$M(x,y) = \sum_{j=1}^{N_\lambda} X_j(x-l_j, y)\, C(x-l_j, y)$$

Where $M = N_x \times (N_y + N_\lambda - 1)$ size coded snapshot image

$X = N_x \times N_y \times N_\lambda$ sized hyperspectral datacube

C = coded aperture pattern of size $N_x \times N_y$ (due to cardboard piece)

l_j = shift in the image of the j^{th} wavelength (due to the prism)

(x,y) = spatial location in coded snapshot.

In multi-frame CASSI,

$$M_t(x,y) = \sum_{j=1}^{N_\lambda} X(x-l_j, y)\, C_t(x-l_j, y) \qquad 1 \le t \le T$$

where

C_t = coded aperture in t^{th} position due to shifting by piezoelectric mechanism.

Q6. $A \rightarrow m \times n$.

We would like $J(A; j, k, l) :=$

$\min_{\alpha_{jkl}} \| A_j - A_{kl} \alpha_{jkl} \|^2$ to have max.

value for any j, k, l where

$A_{kl} \longrightarrow m \times 2$ and $\alpha_{jkl} \longrightarrow 2 \times 1$

We need it to have maximum possible value because we would like any A_j to NOT be well represented as a linear combination of any other two column vectors of A.

The α_{jkl} that minimizes the above expression is given by

$$\alpha_{jkl} = (A_{kl}^T A_{kl})^{-1} A_{kl}^T A_j$$

$$\therefore J(A; jkl) = \| A_j - A_{kl} (A_{kl}^T A_{kl})^{-1} A_{kl}^T A_j \|^2$$

The D.C. criterion is given as

$$\sum_{j, k \neq j} \sum_{\substack{l \neq k \\ l \neq j}} \| A_j - A_{kl} (A_{kl}^T A_{kl})^{-1} A_{kl}^T A_j \|^2$$

OR

$$J(A;j,k,l)$$

$$= \max_{\substack{j,k \neq j, \\ l \neq j, l \neq k}} \| A_j - A_{kl} (A_{kl}^T A_{kl})^{-1} A_{kl}^T A_j \|^2$$

Either of these expressions is fine. In both cases, you need some additional constraints on A, otherwise all elements of A could become ∞ in value. These constraints could be any of the following:

- unit norm constraint on the column or row vectors of A
- $\forall ij \; A_{ij} \in \{0,1\}$ (binary)
- $\forall ij, \; A_{ij} \in [0,1]$

Q7 If $\eta_i \sim \text{Unif}(-a, a)$, then we have $\|\eta\|_2 \leq a\sqrt{m}$

This is a good value for epsilon.

If $\eta_i \sim N(0, \sigma^2)$, then we know that $|\eta_i| \leq 3\sigma$ with a probability $\sim 99\%$. A good thumb rule is to choose $\epsilon = 3\sigma\sqrt{m}$. The fully rigorous technique is to consider that $\|\eta\|_2^2$ is a chi-square r.v. with m-degrees of freedom and choose ϵ based on its tail bounds. But I am not expecting that answer.

Q8 The empirical estimate of the prevalence rate $= k/n = p$.
= prob. that a randomly chosen individual is infected

$1-p =$ prob. that a randomly chosen individual is not infected.

$(1-p)^g =$ prob of obtaining a group of g individuals who are group ALL non-infected.

$1-(1-p)^g =$ prob. of obtaining a group of g individuals containing at least one infected individual.

The expected number of group with at least one infected member

$$= \frac{n}{g} \times \left[1-(1-p)^g\right].$$

The expected number of total tests

$$= \frac{n}{g} + \cancel{g} \times \frac{n}{\cancel{g}} \times \left(1-(1-p)^g\right)$$

$$= \frac{n}{g} + n\left[1-(1-p)^g\right].$$

pooled test in round 1 (cost = q)

Total cost in worst case

individual test (cost = 1) in round 2

$$\propto \frac{n}{g} q + kg = C(g)$$

$$C'(g) = 0 = qn \frac{(-1)}{g^2} + k = 0$$

$$\longrightarrow g = \sqrt{qn/k}$$