Assignment 3: CS 754, Advanced Image Processing

Due: 22nd March before 11:55 pm

Remember the honor code while submitting this (and every other) assignment. All members of the group should work on and understand all parts of the assignment. We will adopt a zero-tolerance policy against any violation.

Submission instructions: You should ideally type out all the answers in Word (with the equation editor) or using Latex. In either case, prepare a pdf file. Create a single zip or rar file containing the report, code and sample outputs and name it as follows: A3-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip. (If you are doing the assignment alone, the name of the zip file is A3-IdNumber.zip). Upload the file on moodle BEFORE 11:55 pm on 22nd March. No assignments will be accepted after a cutoff deadline of 10 am on 23rd March. Note that only one student per group should upload their work on moodle. Please preserve a copy of all your work until the end of the semester. If you have difficulties, please do not hesitate to seek help from me.

1. Download the book 'Statistical Learning with Sparsity: The Lasso and Generalizations' from `https://web.stanford.edu/~hastie/StatLearnSparsity_files/SLS_corrected_1.4.16.pdf`, which is the website of one of the authors. (The book is officially available free of cost). In chapter 11, there are theorems which show error bounds on the minimum of the following objective function: $J(\boldsymbol{\beta}) = \dfrac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_N \|\boldsymbol{\beta}\|_1$ where λ_N is a regularization parameter, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown sparse or compressible signal, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}$ is a measurement vector with N values, $\boldsymbol{w}$ is a zero-mean i.i.d. Gaussian noise vector whose each element has standard deviation σ and $\boldsymbol{X} \in \mathbb{R}^{N \times p}$ is a sensing matrix whose every column is unit normalized. This particular estimator (i.e. minimizer of $J(\boldsymbol{x})$ for $\boldsymbol{x}$) is called the LASSO in the statistics literature. The theorems derive a statistical bound on λ also. The main result is theorem 11.1, parts (a) and (b), and its extension in equation (11.15). These are for sparse signals. For weakly sparse or compressible signals, the result in given in equation 11.16.

   Your task is to answer the following questions:

   (a) Give a careful comparison of the bounds in equations 11.15 and 11.16 to Theorem 3 done in class, and derived in the previous assignment. For this, state how the error bounds vary w.r.t. the number of measurements, the signal sparsity, the noise standard deviation and the signal dimension. Comparing equations 11.15 and 11.16 on one hand to Theorem 3, which of these do you think provides a more intuitive result? Explain why. [2 + 2 + 2 + 2 + 2 = 10 points]
   Solutions: For number of measurements m: In theorem 3, the upper bound increases with m (proportional to $\sqrt{m}$) because the term ϵ (the bound on $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$) increases with m. This is counterintuitive. On the other hand, the LASSO bounds decrease with m (inversely proportional to $\sqrt{m}$). 1 point for the behaviour of LASSO, and 1 point for theorem 3.
   Signal sparsity: The LASSO bounds increase with signal sparsity s, i.e. they are proportional to $\sqrt{s}$. In theorem 3, the upper bounds appear to increase, but the sufficient condition $\delta_{2s} < \sqrt{2} - 1$ will become more stringent as s increases and the number of measurements required for the sufficient condition to hold will also increase as it is proportional to s. 1 point for this part for mentioning the behaviour for LASSO and theorem 3. The bounds in theorem 3 apply to both sparse and weakly sparse (i.e. compressible) signals, whereas the LASSO bounds in theorems 11.15 and 11.16 apply to purely sparse signals only. Bounds for weakly sparse signals are available for LASSO in equation 11.24 with R_q defined in equation 11.7. 1 point for this part for mentioning the behaviour for LASSO and theorem 3.
   Noise standard deviation: In both cases, the bounds increase with the noise standard deviation σ. 1

point each for mentioning the behaviour for LASSO and theorem 3.
Signal dimension: The LASSO bounds scale as $\sqrt{\log p}$ whereas the behaviour of the bounds in theorem 3 have a term in the form $\|\boldsymbol{x} - \boldsymbol{x}_s\|_1$, for which the relationship with $\log p$ is unclear. 2 points here.
Both bounds require m to be of order $s \log p$. The LASSO bounds use the restricted eigenvalue condition, whereas the Theorem 3 bounds use the RIP. But the LASSO bounds are more intuitive in their behaviour w.r.t. the number of measurements. 2 points for this. This is somewhat an open-ended question.

(b) Define the restricted eigenvalue condition (the answer's there in the book and you are allowed to read it, but you also need to understand it). [3 points]
Solution: Refer equation 11.10 of the book. A matrix $\boldsymbol{X}$ of size $N \times p$ (where $n < p$) is said to obey the restricted eigenvalue condition (REC) with constant $\gamma > 0$ if for every vector $\boldsymbol{hat\nu} \in \mathcal{C}$, we have $\dfrac{\hat{\boldsymbol{\nu}}^t \boldsymbol{X}^T \boldsymbol{X} \hat{\boldsymbol{\nu}}}{N\|\hat{\boldsymbol{\nu}}\|_2^2} \geq \gamma$. Clearly, this condition cannot be true for all $\hat{\boldsymbol{\nu}} \in \mathbb{R}^p$ because $N < p$ and $\boldsymbol{X}$ has a non-trivial nullspace. Here $\mathcal{C}$ is a constraint set and we are concerned with vector $\hat{\boldsymbol{\nu}} \in \mathcal{C}$, where $\boldsymbol{\nu} := \boldsymbol{\beta^*} - \hat{\boldsymbol{\beta}}$ is the error vector. In defining, the error vector, note that $\hat{\boldsymbol{\beta}}$ is the estimate of the true vector $\boldsymbol{\beta^*}$ obtained by minimizing $\frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$ with $\lambda \geq \|2\boldsymbol{X}^T\boldsymbol{w}/N\|_\infty$. Here $\boldsymbol{w}$ is a noise vector in $\mathbb{R}^n$ where $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}$. The set $\mathcal{C}$ emerges from equations 11.20, 11.21, 11.22 and Lemma 11.1. Marking scheme: 3 points for a complete definition. The constraint set $\mathcal{C}$ also needs to be defined (and stating that it is 'already there in the book' is not sufficient), otherwise 2 points to be deducted.

(c) Starting from equation 11.20 on page 309 - explain why $G(\hat{\nu}) \leq G(0)$. [3 points] **Solution:** The true signal is β^* and the estimate $\hat{\beta}$ is the minimum of $G(\hat{\nu})$ where $\hat{\nu} = \hat{\beta} - \beta^*$. Since $\hat{\beta}$ is the minimum, we clearly must have $G(\hat{\nu}) = \dfrac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 + \lambda_N\|\hat{\boldsymbol{\beta}}\|_1 \leq \dfrac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta^*}\|^2 + \lambda_N\|\boldsymbol{\beta^*}\|_1 = G(0)$. Marking scheme: 2 points for correct reasoning.

(d) Derive equation 11.21 starting from equation 11.20. [3 points]
Solution: We have $y = X\beta^* + w$. Hence from $G(\hat{v}) \leq G(0)$, we have $\dfrac{1}{2N}\|w - X\hat{\nu}\|^2 + \lambda_N\|\beta^* + \hat{\nu}\|_1 \leq \dfrac{1}{2N}\|w\|^2 + \lambda_N\|\beta^*\|_1$. Opening out the brackets, $\dfrac{1}{2N}\|X\hat{\nu}\|^2 \leq \dfrac{1}{N}w^t X\hat{\nu} + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{\nu}\|_1)$ which is eqn. 11.21. Marking scheme: 2 points for the complete set of steps.

(e) We will now prove theorem 11.2. For this, provide justification for the five equations in the proof of theorem 11.2, parts (a) and (b). [5 points]
Solution: We have $G(\hat{\nu}) \leq G(0)$ as argued earlier which gives rise to the inequality in equation 11.21, i.e.

$$\frac{\|X\hat{\nu}\|_2^2}{2N} \leq \frac{w^t X\nu}{N} + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{\nu}\|_1). \tag{1}$$

The LHS is lower bounded by 0. Moreover, Holder's inequality gives $w^t X\hat{\nu} \leq \|X^T w\|_\infty \|\hat{\nu}\|_1$. This justifies the first relation, that is $0 \leq \|X^T w\|_\infty/N\|\hat{\nu}\|_1 + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \nu\|_1)$.
The second step uses the reverse triangle inequality that $\|\beta^* - (-\hat{\nu})\|_1 \geq -\|\beta^*\|_1 + \|\hat{\nu}\|_1$. Hence $-\|\beta^* - (-\hat{\nu})\|_1 \leq \|\beta^*\|_1 - \|\hat{\nu}\|_1$. After this, some simple algebraic rearrangement is required.
The third step uses the fact that $\lambda_N \geq 2\|X^T w\|_\infty/N$.
The fourth step (the last equation before the proof of 11.25(b) begins), is obtained by starting from equation 11.21 as follows:

$$\frac{\|X\hat{\nu}\|_2^2}{2N} \leq \frac{w^t X\nu}{N} + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{\nu}\|_1) \leq \|X^T w\|_\infty/N\|\hat{\nu}\|_1 + \lambda_N(\|\beta^*\|_1 - \|\beta^*\|_1 + \|\hat{\nu}\|_1). \tag{2}$$

The last inequality is due to the reverse triangle inequality that $\|\beta^* - (-\hat{\nu})\|_1 \geq \|\beta^*\|_1 - \|\hat{\nu}\|_1$. The final inequality is as follows:

$$\frac{\|X\hat{\nu}\|_2^2}{2N} \leq \frac{w^t X\nu}{N} + \lambda_N\|\hat{\nu}\|_1 = \|X^T w\|_\infty\|\hat{\nu}\|_1/N + \lambda_N\|\hat{\nu}\|_1 \leq (\lambda_N/2 + \lambda_N)\|\hat{\nu}\|_1 \leq 3\lambda_N/2 \times 4R_1 = 6\lambda_N R_1, \tag{3}$$

which completes the result.

The fifth step is trickier than what is may appear to be. From the previous result we have the following:

$$\frac{\|X\hat{\nu}\|_2^2}{N} \leq 2\left(\|X^T w\|_\infty/N + \lambda_N\right)\|\hat{\nu}\|_1 \tag{4}$$

$$\leq 2(3\lambda_N/2 + \lambda_N)\|\hat{\nu}\|_1 = 3\lambda_N\|\hat{\nu}\|_1. \tag{5}$$

The cone constraint gives us $\|\hat{\nu}_{S^c}\|_1 \leq 3\|\hat{\nu}_S\|_1$, which means $\|\hat{\nu}\|_1 \leq 4\|\hat{\nu}_S\|_1 \leq 4\sqrt{s}\|\hat{\nu}\|_2$. By REC, we have $\|\hat{\nu}\|_2^2 \leq \|X\hat{\nu}\|_2^2/(N\gamma)$. Substituting these into the previous equation, we get

$$\frac{\|X\hat{\nu}\|_2^2}{N} \leq 12\lambda_N\sqrt{s}\|X\hat{\nu}\|_2/\sqrt{N\gamma} \implies \frac{\|X\hat{\nu}\|_2}{\sqrt{N}} \leq 12\lambda_N\sqrt{s}/\sqrt{\gamma}. \tag{6}$$

Squaring both sides yields us the solution. 2.5 points for this part (a) and 2.5 points for part (b). The steps need to be shown very clearly to get the marks.

2. In class, we studied a video compressive sensing architecture from the paper 'Video from a single exposure coded snapshot' published in ICCV 2011 (See `http://www.cs.columbia.edu/CAVE/projects/single_shot_video/`). Such a video camera acquires a 'coded snapshot' E_u in a single exposure time interval u. This coded snapshot is the superposition of the form $E_u = \sum_{t=1}^{T} C_t \cdot F_t$ where F_t is the image of the scene at instant t within the interval u and C_t is a randomly generated binary code at that time instant, which modulates F_t. Note that E_u, F_t and C_t are all 2D arrays. Also, the binary code generation as well as the final summation all occur within the hardware of the camera. Your task here is as follows:

    (a) Read the 'cars' video in the homework folder in MATLAB using the 'mmread' function which has been provided in the homework folder and convert it to grayscale. Extract the first $T = 3$ frames of the video. You may use the following code snippet:
    ```
    A = mmread('cars.avi'); T = 3; for i=1:T, X(:,:,i) = double(rgb2gray(A.frames(i).cdata));
    end; [H,W,T] = size(X);
    ```

    (b) Generate a $H \times W \times T$ random code pattern whose elements lie in $\{0, 1\}$. Compute a coded snapshot using the formula mentioned and add zero mean Gaussian random noise of standard deviation 2 to it. Display the coded snapshot in your report.

    (c) Given the coded snapshot and assuming full knowledge of C_t for all t from 1 to T, your task is to estimate the original video sequence F_t. For this you should rewrite the aforementioned equation in the form $\boldsymbol{Ax} = \boldsymbol{b}$ where $\boldsymbol{x}$ is an unknown vector (vectorized form of the video sequence). Mention clearly what $\boldsymbol{A}$ and $\boldsymbol{b}$ are, in your report.

    (d) You should perform the reconstruction using either the ISTA algorithm or the OMP algorithm (the original paper used OMP). You can re-use your own code from a previous assignment. For computational efficiency, we will do this reconstruction patchwise. Write an equation of the form $\boldsymbol{Ax} = \boldsymbol{b}$ where $\boldsymbol{x}$ represents the ith patch from the video and having size (say) $8 \times 8 \times T$ and mention in your report what $\boldsymbol{A}$ and $\boldsymbol{b}$ stand for. For perform the reconstruction, assume that each 8×8 slice in the patch is sparse or compressible in the 2D-DCT basis.

    (e) Repeat the reconstruction for all overlapping patches and average across the overlapping pixels to yield the final reconstruction. Display the reconstruction and mention the relative mean squared error between reconstructed and original data, in your report as well as in the code.

    (f) Repeat this exercise for $T = 5, T = 7$ and mention the mention the relative mean squared error between reconstructed and original data again.

    (g) **Note: To save time, extract a portion of about 120×240 around the lowermost car in the cars video and work entirely with it. In fact, you can show all your results just on this part. Some sample results are included in the homework folder.**

    (h) Repeat the experiment with any consecutive 5 frames of the 'flame' video from the homework folder. [20 points = 12 points for correct implementation + 4 points for correct expressions for A,b; 4 points for display results correctly.]

Solution and Marking scheme: For part (c), in the forward model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, we have $\boldsymbol{y}$ to be the vectorized version of the coded snapshot. It has size $n^2 \times 1$ where the snapshot image has size $n \times n$. The matrix $\boldsymbol{A}$ is expressed as $\boldsymbol{A} = (\boldsymbol{\Phi_1}|\boldsymbol{\Phi_2}|...|\boldsymbol{\Phi_T})$ and has size $n^2 \times n^2 T$. Here for each $i \in \{1, 2, ..., T\}$, the matrix $\boldsymbol{\Phi_i}$ is diagonal and has size $n^2 \times n^2$. Also, the underlying spatiotemporal image is expressed as a vector $\boldsymbol{f}$ with $n^2 T$ elements. 4 points for correct equations here.

In patch form, we have $\boldsymbol{y_j} = (\boldsymbol{\Phi_{1j}}|\boldsymbol{\Phi_{2j}}|...|\boldsymbol{\Phi_{Tj}})\boldsymbol{f_j}$ where $\boldsymbol{y_j}$ has size $p^2 \times 1$, each matrix $\boldsymbol{\Phi_{ij}}$ is a diagonal matrix of size $p^2 \times p^2$ and the patch $\boldsymbol{f_j}$ has size $p^2 T \times 1$. These equations carry 2 points.

The values of a Gaussian random variable with mean 0 and standard deviation σ lie in the range $[-3\sigma, +3\sigma]$ with 0.99 probability. Hence a reasonable bound for the error for OMP would be $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \leq 3m\sigma^2$ for m measurements. A more rigorous treatment would involve tail bounds of the chi-square distribution with m degrees of freedom. We would have $P(\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \leq \sigma^2(m + 2m(\sqrt{t} + t))) \geq 1 - \exp(-tm)$. See `https://math.stackexchange.com/questions/2864188/chi-squared-distribution-tail-bound` and references therein.
The first solution with $3\sigma^2$ is less rigorous but acceptable and will be awarded full marks. For a basic working OMP/LASSO, we will award 12 points. 4 points for display of appropriate results. For every result not displayed in the report, 2 points will be deducted. If the result is displayed in the code, but not the report, only 1 point will be deducted. For the flame video, the five consecutive frames should show some motion, otherwise two points are to be deducted. The homework folder has the code for this problem.

3. Consider the image 'cryoem.png' in the homework folder. It is a 2D slice of a 3D macromolecule in the well-known EMDB database. Generate N Radon projections of this 2D image at angles drawn uniformly at random from 0 to 360 degrees. Your job is to implement the Laplacian eigenmaps based algorithm for 2D tomographic reconstruction from unknown angles, given these projection vectors. You should test your algorithm for $N \in \{50, 100, 500, 1000, 2000, 5000, 10000\}$. In each case, display the reconstructed image, and compute its RMSE appropriately normalized for rotation. (Note that cryo-EM reconstructions are valid only up to a global rotation, so direct RMSE computation without compensating for the arbitrary rotation is meaningless). You can use the routines `radon`, `imrotate` in MATLAB. [30 points]
Solution: 15 marks for correct implementation of the Laplacian eigenmaps algorithm, which should ignore the trivial eigenvector containing constant values. The code should choose the two eigenvectors corresponding to the smallest two non-zero eigenvalues. The code should also implement Laplacian eigenmaps. For correct implementation of this, there are 14 points. For correct coding of the angle assignments using a uniform distribution and its order statistics, there are 6 points. There are 4 points for RMSE computation based on rotation normalization and 6 points for properly displaying the reconstructions for different N.

4. Let $R_\theta f(\rho)$ be the Radon transform of the image $f(x, y)$ in the direction given by θ for bin index ρ. Let g be a version of f shifted by (x_0, y_0). Then, prove that $R_\theta g(\rho) = R_\theta f(\rho - (x_0, y_0) \cdot (\cos\theta, \sin\theta))$. [8 points]
Answer: $R(g(x - x_0, y - y_0))(\rho, \theta) = \int\int g(x - x_0, y - y_0)\delta(\rho - x\cos\theta - y\sin\theta)dxdy$. A change of variables $x' = x - x_0, y' = y - y_0$ yields $\int\int g(x', y')\delta(\rho - (x' + x_0)\cos\theta - (y' + y_0)\sin\theta)dx'dy' = \int\int g(x', y')\delta(\rho - x_0\cos\theta - y_0\sin\theta - x'\cos\theta - y'\sin\theta)dx'dy' = R(g(x, y))(\rho - x_0\cos\theta - y_0\sin\theta, \theta)$. Marking scheme: 1 point for the definition of the radon transform substituted correctly in this context. Appropriate change in variables: 4 points; final completion of calculations: 3 points

5. Consider two observed particle images Q_1 and Q_2 corresponding to a 3D density map, each in different 3D orientations and 2D shifts. Let Q_1 be obtained by translating a zero-shift particle image P_1 by $(\delta_{x1}, \delta_{y1})$. Let Q_2 be obtained by translating a zero-shift particle image P_2 by $(\delta_{x2}, \delta_{y2})$. Note that Q_1, Q_2 are practically observed, whereas P_1, P_2 are not observed. Let the common line for the particle images P_1, P_2 pass through the origins of their respective coordinate systems at angles θ_1 and θ_2 with respect to their respective X axes. Derive a relationship between $\delta_{x1}, \delta_{y1}, \theta_1, \delta_{x2}, \delta_{y2}, \theta_2$ and some other observable property of the projection images. Explain how you will determine $\delta_{x1}, \delta_{y1}, \delta_{x2}, \delta_{y2}$ using this equation. Explain how you will extend this relationship to determine the shifts $\{(\delta_{xi}, \delta_{yi})\}_{i=1}^N$ of the N different projection images, and mention the number of knowns and unknowns. [5+2+8+3=18 points]
Solution: We have $Q_1(x, y) = P_1(x + \delta_{x1}, y + \delta_{y1})$ and $Q_2(x, y) = P_2(x + \delta_{x2}, y + \delta_{y2})$. Taking Fourier transforms, we have $\hat{Q}_1(u, v) = e^{\iota 2\pi(u\delta_{x1} + v\delta_{y1})}\hat{P}_1(u, v)$ and $\hat{Q}_2(u, v) = e^{\iota 2\pi(u\delta_{x2} + v\delta_{y2})}\hat{P}_2(u, v)$. Let the common

lines in $\hat{P}_1$ and $\hat{P}_2$ make angles of θ_1 and θ_2 with the local u-axes. Then along the common lines, we have $\hat{P}_1(r\cos\theta_1, r\sin\theta_1) = \hat{P}_2(r\cos\theta_2, r\sin\theta_2)$. Hence, we have:

$$e^{-\iota 2\pi(u\delta_{x1}+v\delta_{y1})}\hat{Q}_1(u,v) = e^{-\iota 2\pi(u\delta_{x2}+v\delta_{y2})}\hat{Q}_2(u,v). \tag{7}$$

Let the phase of $\hat{Q}_1(u,v)/\hat{Q}_2(u,v)$ be given by $e^{\iota r\mu_{12}}$ and let the magnitude be $M(u,v)$. This gives rise to the following

$$e^{-\iota 2\pi(u\delta_{x1}+v\delta_{y1})}M(u,v)e^{-\iota r\mu_{12}} = e^{-\iota 2\pi(u\delta_{x2}+v\delta_{y2})}\hat{Q}_2(u,v). \tag{8}$$

Now, let us substitute $u = r\cos\theta_1, v = r\sin theta_1$ for Q_1 and $u = r\cos\theta_2, v = r\sin theta_2$ for Q_2 since the common line lies along θ_1 and θ_2 in the two different images. This yields

$$e^{-\iota 2\pi r(\cos\theta_1\delta_{x1}+\sin\theta_1\delta_{y1})}M(u,v)e^{-\iota r\mu_{12}} = e^{-\iota 2\pi(r\cos\theta_2\delta_{x2}+r\sin\theta_2\delta_{y2})}\hat{Q}_2(u,v). \tag{9}$$

Equating the phases on either sides give us (after dividing by $\iota 2\pi r$):

$$(\delta_{x1}\cos\theta_1 + \delta_{y1}\sin\theta_1) + \mu_{12} = (\delta_{x2}\cos\theta_2 + \delta_{y2}\sin\theta_2) + 2\pi\times g. \tag{10}$$

This is the desired relationship where g is some integer. The values of $\mu_{12}, \theta_1, \theta_2$ can be determined from the data whereas the values of $\delta_{x1}, \delta_{y1}, \delta_{x2}, \delta_{y2}$ are unknown. This part carries 5 points. The derivation should be shown clearly with appropriate substitutions. If the substitution of u and v in terms of r and θ_1, θ_2 is not done, then 2 points are to be deducted. 1 point to be deducted if the g factor is missing.
This is one equation in 4 unknowns, so it is insufficient to obtain $\delta_{x1}, \delta_{y1}, \delta_{x2}, \delta_{y2}$. 2 points here
Now, if we have N particles, there will be $N(N-1)/2$ such equations for every pair of particles. Let the common line for particles $\hat{P}_i, \hat{P}_j$ be at angles θ_i, θ_j with respect to their corresponding u axes, which gives us:

$$(\delta_{xi}\cos\theta_i + \delta_{yi}\sin\theta_i) + \mu_{ij} = (\delta_{xj}\cos\theta_j + \delta_{yj}\sin\theta_j) + 2\pi g_{ij}. \tag{11}$$

Here g_{ij} is an integer. The phase values μ_{ij} are defined similarly to μ_{12}. This gives us $N(N-1)/2$ equations with $4N$ unknowns, i.e. an overdetermined system. [3 points for the number of knowns and unknowns. Setting up the equations carries 8 points. If the factor $2\pi g_{ij}$ is missing, then 2 points are to be deducted.].

Added note (not expected in the solution): The phase unwrapping ambiguity can be solved as follows in practice. While finding the common lines in Fourier space, one can do the following: Let $\hat{P}_1(r\cos\theta_1, r\sin\theta_1)$ and $\hat{P}_2(r\cos\theta_2, r\sin\theta_2)$ be two radial lines in particles P_1 and P_2. While searching for a common line between $\hat{P}_1$ and $\hat{P}_2$, one needs to multiply (say) $\hat{P}_1$ with a phase factor $e^{\iota 2\pi r\mu_{12}}$ so that the correlation between $\hat{P}_1(r\cos\theta_1, r\sin\theta_1)$ and $\hat{P}_2(r\cos\theta_2, r\sin\theta_2)$ is maximized. The putative μ_{12} values can be restricted to lie in the range $[-0.1L, 0.1L]$ for particles of size $L\times L$. Note that this makes use of the assumption that the shift values are quite small which is reasonable in cryo-em.