

Problem Set 2: Noise, Iterative Reconstruction, and Convexity

CS-491 Machine Learning for Medical Imaging
Fall 2026

Assigned: Week 4 **Due:** start of class, Week 7

Instructions. Submit a single PDF containing your derivations, figures, and discussion, together with the script or notebook that generated every figure. Any programming language is acceptable. You may use libraries for functions unless specified otherwise. Problem 1 is computational and builds on your code from Problem Set 1. Problems 2 and 3 are pen-and-paper.

Problem 1 (Multicoil MRI with Correlated Noise). In this problem you will simulate correlated receiver-coil noise, estimate its covariance, and compute a Tikhonov-regularized reconstruction of undersampled multicoil k-space with two iterative solvers. Use a slice $\mathbf{x} \in \mathbb{C}^N$ and its coil sensitivity maps $\mathbf{s}_1, \dots, \mathbf{s}_C$ from the MoDL brain dataset (`dataset.hdf5`, from the download link at <https://github.com/hkaggarwal/modl>), and reuse your `ft`, `ift`, and `gen_uniform_mask` functions from Problem Set 1.

Recall the multicoil forward model from Problem Set 1,

$$\mathbf{y}_c = \mathbf{I}_\Omega \mathbf{F}_N(\mathbf{s}_c \odot \mathbf{x}) + \boldsymbol{\nu}_c, \quad c = 1, \dots, C, \quad \mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\nu}.$$

Receiver coils sit close together and share electronics, so their noise is correlated. At each k-space location k , the vector of coil noise samples $\boldsymbol{\nu}[k] = [\nu_1[k] \ \cdots \ \nu_C[k]]^\top \in \mathbb{C}^C$ is distributed as $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ with a full covariance matrix $\boldsymbol{\Sigma} \in \mathbb{C}^{C \times C}$, independently across k . We write $\mathbf{s}[n] = [s_1[n] \ \cdots \ s_C[n]]^\top \in \mathbb{C}^C$ for the coil sensitivities at pixel n .

- (a) *Synthetic coil covariance.* Following the synthetic-noise experiments in the Supporting Information of Rep2Rep (Janjušević et al., arXiv:2504.17698), draw a matrix $\mathbf{L} \in \mathbb{R}^{C \times C}$ with entries

$$L_{ij} \sim \begin{cases} \mathcal{N}(\sigma_d, \sigma_j^2), & i = j, \\ \mathcal{U}(-\sigma_c/C, \sigma_c/C), & i \neq j, \end{cases}$$

and generate coil noise as $\boldsymbol{\nu}[k] = \mathbf{L}\mathbf{b}[k]$ with $\mathbf{b}[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_C)$. Use $\sigma_d = 0.03$, $\sigma_j = 0.004$, and $\sigma_c = 0.06$ (one fifth of the values used in the paper). Show that $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^H$. Display $|\boldsymbol{\Sigma}|$ and the coil correlation-coefficient matrix $\rho_{ij} = |\Sigma_{ij}| / \sqrt{\Sigma_{ii}\Sigma_{jj}}$ side-by-side with colorbars.

- (b) *What correlated coil noise looks like.* Simulate noise-only, fully-sampled multicoil k-space ($\mathbf{x} = \mathbf{0}$, every k-space location acquired). Display the magnitude of three coil noise images $\mathbf{F}_N^H \boldsymbol{\nu}_c$ and of the coil-combined noise image $\boldsymbol{\xi} = \sum_c \bar{\mathbf{s}}_c \odot \mathbf{F}_N^H \boldsymbol{\nu}_c$ (the zero-filled combination of Problem Set 1). Show that the coil-combined noise at pixel n has standard deviation

$$\sigma[n] = \sqrt{\mathbf{s}[n]^H \boldsymbol{\Sigma} \mathbf{s}[n]},$$

and verify this by comparing the analytic map with an empirical map computed from 100 noise realizations. Repeat for uncorrelated noise of the same average power, $\boldsymbol{\Sigma}_w = \frac{1}{C} \mathbf{tr}(\boldsymbol{\Sigma}) \mathbf{I}_C$. How do the coil correlations change the spatial distribution of the noise? (Recall that $\sum_c |s_c[n]|^2 = 1$ inside the head.)

- (c) *Estimating the coil covariance.* In practice $\boldsymbol{\Sigma}$ is unknown and must be estimated.

- (i) *From a noise scan.* Scanners can record a short acquisition with the RF transmitter off, which yields independent noise samples $\boldsymbol{\nu}[1], \dots, \boldsymbol{\nu}[N_s]$. Compute the sample covariance $\hat{\boldsymbol{\Sigma}} = \frac{1}{N_s} \sum_{k=1}^{N_s} \boldsymbol{\nu}[k] \boldsymbol{\nu}[k]^H$ and plot its relative error $\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F / \|\boldsymbol{\Sigma}\|_F$ against $N_s \in \{2^6, 2^7, \dots, 2^{14}\}$ on log-log axes. What is the slope, and why?

- (ii) *From the image itself.* Simulate noisy fully-sampled k-space and form the coil images $\mathbf{z}_c = \mathbf{F}_N^H \mathbf{y}_c$. Apply the 2D orthogonal Haar wavelet high-high filter to each coil image, keeping one output per 2×2 block,

$$e_c[i, j] = \frac{1}{2} (z_c[2i, 2j] - z_c[2i+1, 2j] - z_c[2i, 2j+1] + z_c[2i+1, 2j+1]),$$

and use the $N/4$ coil vectors $\mathbf{e}[i, j] \in \mathbb{C}^C$ as noise samples in the sample covariance. Report the relative error. Why does this work, and what limits its accuracy?

- (d) *Whitening.* Let $\mathbf{W} = \hat{\Sigma}^{-1/2}$ (the Hermitian inverse square root) using your estimate from (c)(ii), and whiten the data by applying $\mathbf{W}$ to the coil vector at every k-space location, $\tilde{\mathbf{y}}[k] = \mathbf{W} \mathbf{y}[k]$. Show that the whitened noise $\mathbf{W} \boldsymbol{\nu}[k]$ is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_C)$ when $\hat{\Sigma} = \Sigma$, and that the whitened data obey the multicoil model of Problem Set 1 with whitened coil maps $\tilde{\mathbf{s}}[n] = \mathbf{W} \mathbf{s}[n]$, i.e. $\tilde{\mathbf{y}} = \tilde{\mathbf{A}} \mathbf{x} + \tilde{\boldsymbol{\nu}}$. (Hint: write coil mixing and the per-coil Fourier transform as $\mathbf{W} \otimes \mathbf{I}_N$ and $\mathbf{I}_C \otimes \mathbf{F}_N$, and use the mixed-product property of the Kronecker product.)
- (e) *Tikhonov reconstruction.* With the whitened model, consider

$$\hat{\mathbf{x}}_\lambda = \arg \min_{\mathbf{x}} f(\mathbf{x}), \quad f(\mathbf{x}) = \frac{1}{2} \|\tilde{\mathbf{A}} \mathbf{x} - \tilde{\mathbf{y}}\|_2^2 + \frac{\lambda}{2} \|\mathbf{x}\|_2^2.$$

Derive ∇f and the normal equations satisfied by $\hat{\mathbf{x}}_\lambda$. (For complex $\mathbf{x}$, treat f as a function of the real and imaginary parts of $\mathbf{x}$; the gradient has the same form as in the real case, with transposes replaced by conjugate transposes.) Implement $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{A}}^H$ as functions without forming any matrices, and verify them with an adjoint test: for random $\mathbf{u}$ and $\mathbf{v}$, check that $\langle \tilde{\mathbf{A}} \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \tilde{\mathbf{A}}^H \mathbf{v} \rangle$ to machine precision.

- (f) *Gradient descent vs. conjugate gradients.* Implement, without calling library solvers,
- (i) the power method to estimate $L = \lambda_{\max}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}}) + \lambda$, the Lipschitz constant of ∇f (see Problem 2), and gradient descent $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L} \nabla f(\mathbf{x}_k)$;
 - (ii) the conjugate gradient (CG) method for the normal equations.

Use the variable-density mask `gen_uniform_mask(N2, 4, 0.08)` and start from $\mathbf{x}_0 = \mathbf{0}$. First use CG to sweep λ on a log scale and pick the λ^* with the lowest NRMSE. Then, for $\lambda \in \{\lambda^*, \lambda^*/10, \lambda^*/100\}$, plot the relative residual $\|(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \lambda \mathbf{I}) \mathbf{x}_k - \tilde{\mathbf{A}}^H \tilde{\mathbf{y}}\|_2 / \|\tilde{\mathbf{A}}^H \tilde{\mathbf{y}}\|_2$ (log scale) and the NRMSE against iteration k for both methods, and report the number of iterations each method needs to reach a relative residual of 10^{-4} . Explain the trends using the condition number $\kappa \leq (\lambda_{\max}(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}}) + \lambda) / \lambda$, recalling that per iteration the error of gradient descent shrinks by a factor of at most $(\kappa - 1) / (\kappa + 1)$ and that of CG by roughly $(\sqrt{\kappa} - 1) / (\sqrt{\kappa} + 1)$. For small λ , does the NRMSE decrease monotonically with k ? Why or why not? Display $|\hat{\mathbf{x}}_{\lambda^*}|$ next to $|\mathbf{x}|$ and the zero-filled reconstruction.

- (g) *g-factor.* Undersampling by a factor R costs at least a factor $\sqrt{R}$ in SNR (fewer samples), and the reconstruction amplifies the noise further by the geometry factor $g[n]$,

$$\sigma_R[n] = g[n] \sqrt{R} \sigma_1[n],$$

where $\sigma_R[n]$ is the per-pixel noise standard deviation of the reconstruction from R -fold undersampled data and $\sigma_1[n]$ that of the fully-sampled reconstruction with the same λ . Because $\hat{\mathbf{x}}_\lambda$ is linear in $\tilde{\mathbf{y}}$, you can estimate both by Monte Carlo (“pseudo-replicas”): reconstruct $K \geq 50$ independent realizations of the masked, whitened noise $\tilde{\boldsymbol{\nu}}$ alone and compute the per-pixel standard deviation. Display $g[n]$ inside the head for (i) uniform subsampling with $R = 2$ and $R = 4$ (no ACS) at $\lambda = \lambda^*/100$, and (ii) the variable-density

mask of (f) at λ^* , taking $R = N_2/(\text{number of kept lines})$. Where is g largest, and why? (Hint: which pixels alias onto each other, and how different are their coil sensitivities?) How does λ change g , and what does that cost?

Problem 2 (Convexity, the Descent Lemma, and Majorization-Minimization). Problem 1 used gradient descent with the step $1/L$. This problem derives that algorithm instead of postulating it: each gradient step minimizes a simple quadratic that lies above f , and that alone guarantees the objective goes down.

(a) *Convex functions.* A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex if

$$f(\theta \mathbf{x} + (1 - \theta)\mathbf{z}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{z}) \quad \text{for all } \mathbf{x}, \mathbf{z} \text{ and } \theta \in [0, 1].$$

For a convex function, every local minimum is a global minimum.

- (i) Prove that if $f_1, \dots, f_m$ are convex and $w_1, \dots, w_m \geq 0$, then $\sum_i w_i f_i$ is convex.
 - (ii) Prove that if f is convex, then $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex for any matrix $\mathbf{A}$ and vector $\mathbf{b}$.
 - (iii) Use (i), (ii), and the second-derivative test for functions of one variable to show that the Tikhonov objective of Problem 1 is convex. Then show, directly from the definition and the properties of a norm, that every norm $\|\cdot\|$ is convex.
- (b) *The Descent Lemma and majorization-minimization.* Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be differentiable with an L -Lipschitz gradient,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{z})\|_2 \leq L\|\mathbf{x} - \mathbf{z}\|_2 \quad \text{for all } \mathbf{x}, \mathbf{z}.$$

We first show that f never rises faster than a quadratic away from any point, then turn that bound into an algorithm.

- (i) Fix $\mathbf{x}$ and $\mathbf{z}$, and let $\varphi(t) = f(\mathbf{x} + t(\mathbf{z} - \mathbf{x}))$ for $t \in [0, 1]$. This is f restricted to the line segment from $\mathbf{x}$ ($t = 0$) to $\mathbf{z}$ ($t = 1$). Use the chain rule to show that $\varphi'(t) = \nabla f(\mathbf{x} + t(\mathbf{z} - \mathbf{x}))^\top (\mathbf{z} - \mathbf{x})$.
- (ii) The fundamental theorem of calculus gives $f(\mathbf{z}) - f(\mathbf{x}) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt$. Subtract $\nabla f(\mathbf{x})^\top (\mathbf{z} - \mathbf{x})$ from both sides, writing it as $\int_0^1 \nabla f(\mathbf{x})^\top (\mathbf{z} - \mathbf{x}) dt$, to show that

$$f(\mathbf{z}) - \left(f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{z} - \mathbf{x}) \right) = \int_0^1 (\nabla f(\mathbf{x} + t(\mathbf{z} - \mathbf{x})) - \nabla f(\mathbf{x}))^\top (\mathbf{z} - \mathbf{x}) dt.$$

The left-hand side is the error of the first-order Taylor approximation of f around $\mathbf{x}$.

- (iii) Bound the integrand in two steps: the Cauchy–Schwarz inequality $\mathbf{u}^\top \mathbf{v} \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$, then the Lipschitz condition applied to the points $\mathbf{x} + t(\mathbf{z} - \mathbf{x})$ and $\mathbf{x}$. Integrate the result, using $\int_0^1 t dt = \frac{1}{2}$, to prove the Descent Lemma,

$$f(\mathbf{z}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{z} - \mathbf{x}) + \frac{L}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 \quad \text{for all } \mathbf{x}, \mathbf{z}.$$

- (iv) *A majorizer.* For a fixed point $\mathbf{x}$, define the quadratic

$$g(\mathbf{z}; \mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{z} - \mathbf{x}) + \frac{L}{2} \|\mathbf{z} - \mathbf{x}\|_2^2.$$

Show that $g(\cdot; \mathbf{x})$ majorizes f at $\mathbf{x}$: $g(\mathbf{z}; \mathbf{x}) \geq f(\mathbf{z})$ for every $\mathbf{z}$, and $g(\mathbf{x}; \mathbf{x}) = f(\mathbf{x})$. As an example, $f(z) = \sqrt{1 + z^2}$ has $f''(z) = (1 + z^2)^{-3/2} \leq 1$, so $L = 1$. Sketch f and $g(\cdot; 2)$ on the same axes.

(v) *Minimizing the majorizer.* Complete the square to show that

$$g(\mathbf{z}; \mathbf{x}) = f(\mathbf{x}) - \frac{1}{2L} \|\nabla f(\mathbf{x})\|_2^2 + \frac{L}{2} \left\| \mathbf{z} - \left(\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x}) \right) \right\|_2^2,$$

and conclude that the minimizer of $g(\cdot; \mathbf{x})$ is $\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x})$: one step of gradient descent with step size $1/L$. Mark this point on your sketch from (iv).

(vi) *Majorization-minimization (MM).* The MM algorithm repeatedly minimizes the current majorizer, $\mathbf{x}_{k+1} = \arg \min_{\mathbf{z}} g(\mathbf{z}; \mathbf{x}_k)$, which by (v) is gradient descent. Justify each step of the chain

$$f(\mathbf{x}_{k+1}) \leq g(\mathbf{x}_{k+1}; \mathbf{x}_k) \leq g(\mathbf{x}_k; \mathbf{x}_k) = f(\mathbf{x}_k),$$

so the objective never increases. Then use (v) to show the sharper statement

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L} \|\nabla f(\mathbf{x}_k)\|_2^2,$$

and sum it over $k = 0, \dots, K-1$ to show

$$\min_{0 \leq k < K} \|\nabla f(\mathbf{x}_k)\|_2^2 \leq \frac{2L(f(\mathbf{x}_0) - f^*)}{K},$$

where f^* is the minimum value of f . The gradient therefore becomes arbitrarily small.

(vii) *The step size in Problem 1.* Show that the gradient of $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \|\mathbf{x}\|_2^2$ is Lipschitz with $L = \lambda_{\max}(\mathbf{A}^\top \mathbf{A}) + \lambda$. (Hint: write $\nabla f(\mathbf{x}) - \nabla f(\mathbf{z})$ as a matrix times $\mathbf{x} - \mathbf{z}$, and recall that the spectral norm of a symmetric positive semidefinite matrix is its largest eigenvalue.) What would go wrong in MM if you used a smaller constant than L in g ?

Problem 3 (Dual Norms, Conjugates, and Lagrange Multipliers). Duality pairs each norm with a dual norm, each function with its conjugate, and each constraint with a Lagrange multiplier. All vectors below are real.

(a) *Dual norms and Fenchel conjugates.* The dual norm of a norm $\|\cdot\|$ and the Fenchel conjugate of a function f are

$$\|\mathbf{z}\|_* = \sup_{\|\mathbf{x}\| \leq 1} \mathbf{z}^\top \mathbf{x}, \quad f^*(\mathbf{z}) = \sup_{\mathbf{x}} (\mathbf{z}^\top \mathbf{x} - f(\mathbf{x})).$$

- (i) Show that the dual norm of $\|\cdot\|_1$ is $\|\cdot\|_\infty$: prove the inequality $\mathbf{z}^\top \mathbf{x} \leq \|\mathbf{z}\|_\infty \|\mathbf{x}\|_1$, then find an $\mathbf{x}$ with $\|\mathbf{x}\|_1 = 1$ that achieves equality.
 - (ii) Show that f^* is convex for every f , convex or not. (Hint: f^* is a pointwise supremum of functions of $\mathbf{z}$. What kind of functions?)
 - (iii) Compute f^* for $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_2^2$, and for $f(\mathbf{x}) = \|\mathbf{x}\|$, an arbitrary norm. Express the second answer in terms of the dual norm.
- (b) *Lagrange multipliers.* A constraint can be moved into the objective by adding it with a weight, the Lagrange multiplier. To minimize $f(\mathbf{x})$ subject to equality constraints $\mathbf{h}(\mathbf{x}) = \mathbf{0}$, form the *Lagrangian*

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}) = f(\mathbf{x}) + \boldsymbol{\nu}^\top \mathbf{h}(\mathbf{x}).$$

At a minimizer $\mathbf{x}^*$ there is a multiplier $\boldsymbol{\nu}^*$ such that $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*) = \mathbf{0}$ (stationarity) and $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$ (feasibility). Geometrically, stationarity says that at the optimum ∇f is a combination of the constraint gradients: moving along the constraint set cannot decrease f to first order. For an inequality constraint $g(\mathbf{x}) \leq 0$, the Lagrangian is $f(\mathbf{x}) + \mu g(\mathbf{x})$ with $\mu \geq 0$, and the conditions, called the Karush–Kuhn–Tucker (KKT) conditions, are:

- stationarity, $\nabla f(\mathbf{x}^*) + \mu^* \nabla g(\mathbf{x}^*) = \mathbf{0}$;
- primal feasibility, $g(\mathbf{x}^*) \leq 0$;
- dual feasibility, $\mu^* \geq 0$;
- complementary slackness, $\mu^* g(\mathbf{x}^*) = 0$: either the constraint is active ($g(\mathbf{x}^*) = 0$), or $\mu^* = 0$ and the constraint plays no role.

For the convex problems below, these conditions are both necessary and sufficient for optimality.

- (i) *Warm-up.* Minimize $x_1^2 + x_2^2$ subject to $x_1 + x_2 = 1$. Write the Lagrangian with a multiplier ν , set $\partial \mathcal{L} / \partial x_1 = 0$ and $\partial \mathcal{L} / \partial x_2 = 0$, and solve these two equations together with the constraint for x_1 , x_2 , and ν . On a sketch of the circular level sets of f and the constraint line, draw ∇f and the constraint gradient at the solution, and check that they are parallel.
- (ii) *Minimum-norm solutions.* Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m < n$ and full row rank, so that $\mathbf{A}\mathbf{x} = \mathbf{y}$ has infinitely many solutions. To find the one of smallest norm, solve

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}$$

in steps:

- (1) write the Lagrangian with a multiplier $\boldsymbol{\nu} \in \mathbb{R}^m$;
 - (2) use stationarity to express $\mathbf{x}$ in terms of $\boldsymbol{\nu}$;
 - (3) substitute into the constraint, explain why $\mathbf{A}\mathbf{A}^\top$ is invertible, and solve for $\boldsymbol{\nu}$;
 - (4) conclude that $\mathbf{x}^* = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{y}$, and use the SVD $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ to show that this equals the pseudo-inverse solution $\mathbf{A}^\dagger \mathbf{y}$ of Problem Set 1.
- (iii) *Choosing λ from the noise level.* In Problem 1 we picked λ using the ground truth, which is unavailable in practice. Suppose instead we ask for the smallest-norm image that fits the data only as well as the noise allows:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \frac{1}{2} (\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 - \epsilon) \leq 0,$$

with $0 < \epsilon < \|\mathbf{y}\|_2^2$. Work through the KKT conditions:

- (1) write the Lagrangian with a multiplier $\mu \geq 0$;
- (2) show that stationarity reads $(\mathbf{I} + \mu \mathbf{A}^\top \mathbf{A})\mathbf{x} = \mu \mathbf{A}^\top \mathbf{y}$;
- (3) show that $\mu = 0$ leads to an infeasible point, so $\mu > 0$, and that dividing by μ gives the Tikhonov normal equations with $\lambda = 1/\mu$;
- (4) use complementary slackness to show that the solution satisfies $\|\mathbf{A}\mathbf{x}_\lambda - \mathbf{y}\|_2^2 = \epsilon$.

So the right λ is the one whose residual equals ϵ . If $\mathbf{y} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{n}$ with noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$, compute $\mathbb{E} \|\mathbf{A}\mathbf{x}_{\text{true}} - \mathbf{y}\|_2^2$ and use it to choose ϵ . After the whitening of Problem 1(d), what is ϵ in terms of C and $|\Omega|$? (The residual increases with λ , so the matching λ can be found by bisection. Optionally, try it in Problem 1 and compare with λ^* .)