Christian Kent

**Math 290 Problem Set 2 Due: Tuesday 3/31/2026 at 11:59 p.m. EST**

**Instruction**: Read the assignment policy. For problems 5 and 9, include a printout your code with your homework submission. You should submit your assignment on Gradescope.

**1**. Consider a linear autoencoder with encoder $\mathcal{E}(\mathbf{x}) = \mathbf{W}_E\mathbf{x}$, and decoder $\mathcal{D}(\mathbf{z}) = \mathbf{W}_D\mathbf{z}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ be a collection of data points, and let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ denote the data matrix. Assume the latent dimension is $k < d$, so that $\mathbf{W}_E \in \mathbb{R}^{k\times d}$ and $\mathbf{W}_D \in \mathbb{R}^{d\times k}$. We consider the following reconstruction error:

$$\min_{\mathbf{W}_E, \mathbf{W}_D} \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i - \mathcal{D}(\mathcal{E}(\mathbf{x}_i))\|_2^2 = \min_{\mathbf{W}_E, \mathbf{W}_D} \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{W}_D\mathbf{W}_E\mathbf{x}_i\|_2^2 .$$

(a) Show that the optimal matrix $\mathbf{W}_D\mathbf{W}_E$ is the orthogonal projection onto the subspace spanned by the top $k$ left singular vectors of $\mathbf{X}$.

(b) Provide one explicit solution for $\mathbf{W}_D$ and $\mathbf{W}_E$, and argue why this solution is not unique.

(c) What is the solution if we additionally require $\mathbf{W}_D$ and $\mathbf{W}_E^\top$ to have orthonormal columns?

(d) Explain how your conclusions in parts (a)–(c) relate to principal component analysis (PCA).

**2**. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ be unit vectors representing $n$ distinct lines in $\mathbb{R}^d$. Assume the lines are equiangular, i.e., there exists $\alpha \in (0, 1)$ such that

$$|\langle \mathbf{x}_i, \mathbf{x}_j\rangle| = \alpha \qquad \text{for all } i \neq j.$$

(a) For each $i$, define the rank-one projection $\mathbf{P}_i := \mathbf{x}_i\mathbf{x}_i^\top \in \mathrm{Sym}_d$, where $\mathrm{Sym}_d$ denotes the space of real symmetric $d \times d$ matrices. Show that

$$\langle \mathbf{P}_i, \mathbf{P}_j\rangle_F = \begin{cases} 1, & i = j, \\ \alpha^2, & i \neq j, \end{cases}$$

where $\langle \mathbf{A}, \mathbf{B}\rangle_F = \mathrm{tr}(\mathbf{A}^\top\mathbf{B})$ is the Frobenius inner product.

(b) Let $\mathbf{G}$ be the Gram matrix of $\{\mathbf{P}_1, \ldots, \mathbf{P}_n\}$ with respect to the Frobenius inner product. Write $\mathbf{G}$ explicitly and compute its eigenvalues. Deduce that $\{\mathbf{P}_i\}_{i=1}^n$ is a linearly independent set in $\mathrm{Sym}_d$.

(c) Use the fact that $\dim(\mathrm{Sym}_d) = \frac{d(d+1)}{2}$ to conclude that

$$n \leq \frac{d(d+1)}{2}.$$

Do some research to find the name of this bound.

(d) Suppose that equality holds, so that $\{\mathbf{P}_i\}$ spans $\mathrm{Sym}_d$. Explain why it is impossible to add an additional rank-one projection $\mathbf{y}\mathbf{y}^\top$ that is equiangular with all $\mathbf{P}_i$.

(e) Briefly argue that an equiangular set of unit vectors achieving equality in the Welch bound must satisfy

$$n \leq \frac{d(d+1)}{2},$$

**3**. This problem is concerned with applying algorithm unrolling to structured dictionary learning problem. Assume we have $n$ data points $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$ in $\mathcal{R}^d$. We consider the following structured dictionary learning problem. For a data point $\mathbf{y} \in \mathbb{R}^d$, we consider the optimization problem

$$\min_{\mathbf{x} \in \Delta^m} \ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \ + \ \lambda \sum_{j=1}^{m} x_j \|\mathbf{y} - \mathbf{a}_j\|_2^2,$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m] \in \mathbb{R}^{d \times m}$ is a learnable dictionary, $\mathbf{x} \in \mathbb{R}^m$ is constrained to the probability simplex $\Delta^m := \{\mathbf{x} \in \mathbb{R}^m : x \geq 0, \ \mathbf{1}^\top \mathbf{x} = 1\}$, and $\lambda > 0$ is a regularization parameter. Given the data points $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$, the goal is to learn both the dictionary $\mathbf{A}$ and the corresponding sparse coefficients $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$.

(a) Provide a brief interpretation of the dictionary learning problem. What kind of structure do we impose on the representation of a data point using the atoms? What happens as $\lambda \to \infty$?

(b) Let $\mathbb{T} := \{\mathbf{r} \in \mathbb{R}^m : \sum_{i=1}^m r_i = 1\}$. Show that the orthogonal projection of a vector $\mathbf{z} \in \mathbb{R}^m$ onto $\mathbb{T}$ has the form

$$\tilde{z}_i = z_i - \tau,$$

and derive the value of $\tau$. Provide a geometric explanation for why the projection direction must lie along the normal vector of $\mathbb{T}$.

(c) Consider the Euclidean projection $\Pi_{\Delta^m}(\mathbf{z}) = \arg\min_{\mathbf{p} \in \Delta^m} \frac{1}{2}\|\mathbf{p} - \mathbf{z}\|_2^2$.. Using the intuition from the affine projection in Part (b), argue that the solution must take the form $\max(z_i - \tau, 0)$.

(d) Assume the entries of $\mathbf{z}$ are sorted in decreasing order and define the active set

$$S := \{i : z_i - \tau > 0\}.$$

Show that

$$\tau = \frac{\sum_{i \in S} z_i - 1}{|S|}.$$

(e) **Bonus**: Write down the KKT conditions for the simplex projection problem and explain how they lead to the thresholding structure above.

(f) Briefly compare the simplex projection to the Softmax mapping. What sparsity-inducing property does the simplex projection possess?

(g) Fix the dictionary $\mathbf{A}$. Define

$$f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \sum_{j=1}^{m} x_j \|\mathbf{y} - \mathbf{a}_j\|_2^2.$$

Show that $f(\mathbf{x})$ is convex in $\mathbf{x}$, and compute the gradient $\nabla_{\mathbf{x}} f(\mathbf{x})$.

(h) Write the projected gradient descent update for minimizing $f(\mathbf{x})$ over $\Delta^m$:

$$\mathbf{x}^{(k+1)} = \Pi_{\Delta^m}\left(\mathbf{x}^{(k)} - \eta \nabla_x f(\mathbf{x}^{(k)})\right),$$

where $\eta > 0$ is a step size. Assume we run $T$ iterations of the projected gradient descent algorithm starting from an initialization $\mathbf{x}^{(0)}$. Use algorithm unrolling to design an autoencoder that would learn $\mathbf{A}$ given training data. You can assume $\eta$ is a fixed parameter.

**4**. Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m] \in \mathbb{R}^{d\times m}$ have unit-norm columns, i.e. $\|\mathbf{a}_j\|_2 = 1$ for all $j$. For an index set $S \subseteq [m]$ with $|S| = s$, let $\mathbf{A}_S \in \mathbb{R}^{d\times s}$ denote the submatrix containing the columns of $\mathbf{A}$ indexed by $S$.

**Definition (RIP).** We say that $\mathbf{A}$ satisfies the *restricted isometry property* (RIP) of order $s$ with constant $\delta_s \in (0,1)$ if for every $s$-sparse vector $\mathbf{x} \in \mathbb{R}^m$,

$$(1-\delta_s)\|\mathbf{x}\|_2^2 \;\le\; \|\mathbf{A}\mathbf{x}\|_2^2 \;\le\; (1+\delta_s)\|\mathbf{x}\|_2^2. \tag{1}$$

(a) Fix any set $S \subseteq [m]$ with $|S| = s$. Show that RIP of order $s$ implies that for all $u \in \mathbb{R}^s$,

$$(1-\delta_s)\|\mathbf{u}\|_2^2 \;\le\; \|\mathbf{A}_S\mathbf{u}\|_2^2 \;\le\; (1+\delta_s)\|\mathbf{u}\|_2^2.$$

(b) Deduce that the Gram matrix $\mathbf{G}_S := \mathbf{A}_S^\top \mathbf{A}_S$ satisfies the bounds

$$(1-\delta_s)\mathbf{I}_s \;\preceq\; \mathbf{G}_S \;\preceq\; (1+\delta_s)\mathbf{I}_s,$$

where $\mathbf{I}_s$ is an $s \times s$ identity matrix. Conclude that all eigenvalues of $\mathbf{G}_S$ lie in the interval $[\,1-\delta_s,\; 1+\delta_s\,]$.

(c) Using (b), explain why $\mathbf{A}_S^\top \mathbf{A}_S$ is invertible whenever $\delta_s < 1$. Moreover, show that

$$\|(\mathbf{A}_S^\top \mathbf{A}_S)^{-1}\|_2 \;\le\; \frac{1}{1-\delta_s}.$$

(d) Assume $\mathbf{A}$ satisfies RIP of order $s \ge 2$ with constant $\delta_s$. Show that for any $i \ne j$,

$$|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \;\le\; \delta_s,$$

and conclude that $\mu(\mathbf{A}) \le \delta_s$.

**5**. In this problem you will experiment with an autoencoder inspired by the model discussed in Problem 3. The first step is to generate a dataset as follows:

$$t \sim \mathrm{Unif}(t_{\min}, t_{\max}), \qquad y(t) = \begin{bmatrix} t\cos t \\ t \sin t \end{bmatrix} + \sigma \boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2).$$

Set $t_{\min} = 1, t_{\max} = 10.0$ and $\sigma = 0.05$. This produces data points $y_i \in \mathbb{R}^2$ lying near a one-dimensional manifold. Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m] \in \mathbb{R}^{2\times m}$ be a learnable dictionary. For each data point $\mathbf{y}_i$, we infer a latent code $\mathbf{x}_i \in \Delta^m$ by minimizing

$$\min_{\mathbf{x}_i \in \Delta^m} \; \|\mathbf{y}_i - \mathbf{A}\mathbf{x}_i\|_2^2 \;+\; \lambda \sum_{j=1}^{m} (x_i)_j \, \|\mathbf{y}_i - \mathbf{a}_j\|_2^2,$$

where $\Delta^m$ denotes the probability simplex.

**Inference (Encoder).** For fixed $\mathbf{A}$, infer $\mathbf{x}_i$ using projected gradient descent (PGD):

$$\mathbf{x}_i^{(k+1)} = \Pi_{\Delta^m}\Big(\mathbf{x}_i^{(k)} - \eta\left[2\mathbf{A}^\top(\mathbf{A}\mathbf{x}_i^{(k)} - \mathbf{y}_i) + \lambda\, d(\mathbf{y}_i, \mathbf{A})\right]\Big),$$

where $d(\mathbf{y}_i, \mathbf{A}) \in \mathbb{R}^m$ has entries

$$d_j(\mathbf{y}_i, \mathbf{A}) = \|\mathbf{y}_i - \mathbf{a}_j\|_2^2.$$

**Dictionary Update (Decoder).** For fixed codes $\{\mathbf{x}_i\}_{i=1}^N$, update $\mathbf{A}$ by solving the linear system

$$\mathbf{A}\big(\mathbf{X}^\top\mathbf{X} + \lambda\,\mathrm{diag}(\mathbf{s}) + \beta\mathbf{I}\big) = (1+\lambda)\mathbf{Y}^\top\mathbf{X}, \qquad \mathbf{s} := \mathbf{X}^\top\mathbf{1},$$

where $\mathbf{X} \in \mathbb{R}^{N\times m}$ stacks the codes and $\mathbf{Y} \in \mathbb{R}^{N\times 2}$ stacks the data.

(a) Generate $N = 2000$ data points from the 2D swiss-roll model and plot them.

(b) Implement Euclidean projection onto the probability simplex $\Delta^m$.

(c) Implement PGD-based inference for the latent codes $\mathbf{x}_i$.

(d) Learn the dictionary $\mathbf{A}$ by alternating between sparse coding and dictionary learning.

(e) Plot the learned atoms $\mathbf{a}_j$ overlaid on the data.

(f) Report the final average objective value and the average support size of the latent codes $\#\{j : x_{ij} > 10^{-3}\}$.

[**Remark**: You can experiment with different parameters. Here is one choice: $m = 16$, $\lambda = 10$, $T = 30, \eta = 0.2$, $T = 25$, $\beta = 10^{-2}$, where $T$ denotes the total number of iterations.]

**6**. Given a set of points $\mathbf{x}_1, ..., \mathbf{x}_k$ in $\mathcal{R}^d$, and a similarity or weight matrix $\mathbf{W}$, the Laplacian is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ denotes the degree matrix. Assume that all diagonal entries of $\mathbf{D}$ are non-zero. Consider the optimization problem

$$\min_{\mathbf{Y} \in \mathbb{R}^{k \times m}} \quad \text{trace}(\mathbf{Y}^\top \mathbf{L} \mathbf{Y}),$$

where the rows of $\mathbf{Y}$ represent the $m$-dimensional embedding of $k$ points.

(a) What is the importance of the constraint $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$ i.e., what is the solution obtained without this constraint?

(b) Derive the Lagrangian for the constrained optimization problem

$$\min_{\mathbf{Y}} \ \text{trace}(\mathbf{Y}^\top \mathbf{L} \mathbf{Y}) \quad \text{subject to} \quad \mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_m.$$

(c) Show that any optimal solution $\mathbf{Y}$ consists of eigenvectors of $\mathbf{L}$. Which eigenvectors should be chosen to minimize the objective?

(d) Consider the alternative constraint $\mathbf{Y}^\top \mathbf{D} \mathbf{Y} = \mathbf{I}_m$. Derive the optimality conditions for

$$\min_{\mathbf{Y}} \ \text{trace}(\mathbf{Y}^\top \mathbf{L} \mathbf{Y}) \quad \text{subject to} \quad \mathbf{Y}^\top \mathbf{D} \mathbf{Y} = \mathbf{I}_m.$$

Show that the columns of an optimal solution $\mathbf{Y}$ satisfy the generalized eigenvalue problem $\mathbf{L}\mathbf{y} = \lambda \mathbf{D}\mathbf{y}$.

(e) Explain which generalized eigenvectors should be chosen to form $\mathbf{Y}$, and how this choice depends on the embedding dimension $m$.

(f) **Exploration:** Do some research and briefly discuss how this finite-dimensional problem relates to eigenfunction problems of the Laplace–Beltrami operator in the continuous setting.

(g) **Bonus**: Suppose we add the constraint $\mathbf{Y}^\top \mathbf{D} \mathbf{1} = \mathbf{0}$ to the optimization problem in part (d). Explain how this affects the optimal solution.

**7**. Find all the eigenvalues and characterize the corresponding eigenvectors for the Laplacian of the complete graph.

**8**. The quadratic form of the Laplacian is $\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{(i,j)\in E} A_{ij}(x_i - x_j)^2$. Consider an eigenvector $\mathbf{v}$ corresponding to the second smallest eigenvalue $\lambda_2$. Given that $\mathbf{v}$ must be orthogonal to the constant vector ($\mathbf{v} \perp \mathbf{1}$), use the interpretation of the quadratic form to argue which graph would have a larger $\lambda_2$: a Path Graph ($P_n$) or a Complete Graph ($K_n$)? A path graph (denoted as $P_n$) is a specific type of graph that consists of a single sequence of vertices where each vertex is connected to the next by exactly one edge.

**9.** The goal of this problem is to explore graph cuts and spectral clustering.

(a) Construct a connected graph of your choice with at least six nodes for which the minimum cut yields a trivial partition into two clusters. Compute the Fiedler vector (the eigenvector corresponding to the second-smallest eigenvalue of the graph Laplacian). Apply thresholding to the Fiedler vector and determine whether it correctly identifies the two clusters.

(b) Repeat the experiment in part (a) using a connected graph with at least nine nodes, where the minimum cut solution corresponding to three clusters is visually or structurally apparent. Compute the eigenvectors corresponding to the first two nonzero smallest eigenvalues of the graph Laplacian. Form an embedding using these eigenvectors and cluster the rows of the resulting eigenvector matrix to obtain three clusters. Does this procedure recover the underlying cluster structure?

**10.** Let $\mathbf{Y} \in \mathbb{R}^{m \times n}$ be a data matrix with $n$ data points, and consider the optimization problem

$$\min_{\mathbf{D} \in \mathbb{R}^{m \times k},\, \mathbf{X} \in \mathbb{R}^{k \times n}} F(\mathbf{D}, \mathbf{X}) := \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_1,$$

where $\lambda > 0$ and $m < k$ (overcomplete dictionary). In Parts (b)–(f), additionally assume the constraint $\mathbf{X} \geq \mathbf{0}$.

(a) Show that $F(\mathbf{D}, \mathbf{X})$ is convex in $\mathbf{X}$ for fixed $\mathbf{D}$, and convex in $\mathbf{D}$ for fixed $\mathbf{X}$.

[**Hint**: For the block-wise convexity, you may use the fact that the composition of a convex function with a linear mapping is convex. you may argue noting that $\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2$ is a quadratic function and compute the Hessian with respect to the two variables].

(b) Explain why the nonnegativity constraint $\mathbf{X} \geq \mathbf{0}$ preserves convexity of the $\mathbf{X}$-subproblem.

(c) Explain why the problem is ill-posed if no constraint is imposed on $\mathbf{D}$.

(d) To address the issue in part (c), suppose that the dictionary is constrained by

$$\mathcal{D} := \{\mathbf{D} \in \mathbb{R}^{m \times k} : \|\boldsymbol{d}_j\|_2 \leq 1 \text{ for all } j\}.$$

Show that the $D$-subproblem remains convex under this constraint, and explain why such a constraint is natural.

(e) Consider the alternating minimization scheme

$$\mathbf{X}^{(t+1)} \in \arg\min_{\mathbf{X} \geq \mathbf{0}} F(\mathbf{D}^{(t)}, \mathbf{X}),$$
$$\mathbf{D}^{(t+1)} \in \arg\min_{\mathbf{D} \in \mathcal{D}} F(\mathbf{D}, \mathbf{X}^{(t+1)}),$$

where each subproblem is solved exactly. Show that the objective values form a nonincreasing sequence.

(f) Using your result from part (e), show that the sequence of objective values

$$F(\mathbf{D}^{(t)}, \mathbf{X}^{(t)})$$

converges. Explain why convergence of the objective values does *not* imply convergence of the iterates

$$(\mathbf{D}^{(t)}, \mathbf{X}^{(t)}).$$

(1) Consider a linear autoencoder w/ encoder $\mathcal{E}(\vec{x}) = W_E \vec{x}$ and decoder $\mathcal{D}(\vec{z}) = W_D \vec{z}$. Let $x_1, \dots, x_n \in \mathbb{R}^d$ be a collection of data w/ $X = [x_1, \dots, x_n] \in \mathbb{R}^{d \times n}$ the data matrix. Assume the latent dimension is $k < d$, so $W_E \in \mathbb{R}^{k \times d}$ and $W_D \in \mathbb{R}^{d \times k}$. We consider the reconstruction error:

$$\min_{W_E, W_D} \frac{1}{n} \sum_{i=1}^{n} \| x_i - \mathcal{D}(\mathcal{E}(x_i)) \|_2^2 = \min_{W_E, W_D} \frac{1}{n} \sum_{i=1}^{n} \| x_i - W_D W_E x_i \|_2^2$$

(a) Show that the optimal soln is the orthogonal projection onto the subspace spanned by the top $k$ left singular vectors.

Pf: This result is rather intuitive since we are minimizing an objective $\|x_i - Ay_i\|_2^2$ where $A$ is a linear map; so we want an optimal orthogonal projection onto $\text{col}(A)$. First, fix $W_D$ and write $W_E x_i = z_i$ to denote the latent codes. Thus, an objective $f : \mathbb{R}^k \to \mathbb{R}$ and an optimization problem

$$\text{for} \quad f(z_i) = \| x_i - W_D z_i \|_2^2 \quad \underline{\text{and}} \quad \min_{z_i \in \mathbb{R}^k} f(z_i) \text{ s.t. } z_i = W_E x_i, \quad (1)$$

we can determine an optimal encoding map $W_E$ given fixed decoding map $W_D$.

This is a standard LQ problem. So, we compute the minimizer as

$$\nabla f(z_i) = W_D^T W_D z_i - W_D^T x_i \implies z_i = (W_D^T W_D)^{-1} W_D^T x_i = W_D^+ x_i.$$

So, together with (1), we see that $W_E x_i = W_D^+ x_i$. Hence, given a fixed decoding map $W_D$, the optimal encoding map $W_E$ is $W_D^+$.

As a sanity check, note that $W_D^+, W_E \in \mathbb{R}^{d \times k}$.

$\therefore$ The autoencoder map is $(\mathcal{D} \circ \mathcal{E})(x) = W_D W_D^+ x$.

As $W_D \in \mathbb{R}^{d \times k}$ for $k << d$, then we can assume that $W_D$ has full column rank, and therefore $W_D^T W_D \in \mathbb{R}^{k \times k}$ is invertible. Thus, $\mathcal{D} \circ \mathcal{E}$ is well-defined in this manner. Furthermore, as

$$[(\mathcal{D} \circ \mathcal{E})] W_D = W_D W_D^+ W_D = W_D$$

by definition of the Moore-Penrose pseudoinverse, then the autoencoder is an orthogonal projection onto $\text{COL}(W_D)$.

Now, it is enough to conclude by showing that $W_D$ is the matrix whose columns contain the top $k$ left singular vectors of data matrix $X$. To do this we will reformulate the objective as

$$\min_{W_E, W_D} \frac{1}{n} \sum_{i=1}^{n} \| x_i - W_D W_E x_i \|_2^2 = \min_{W_D} \frac{1}{n} \sum_{i=1}^{n} \| x_i - W_D W_D^+ x_i \|_2^2. \quad (2)$$

To simplify analysis and notation, assume that $v_1, \dots, v_d \in \mathbb{R}^d$ is an orthonormal basis for $\mathbb{R}^d$, which we can always construct via the Gram-Schmidt Thm.

Let $V$ have $v_1, \dots, v_d$ along its columns. We will write

$$x_i = \sum_{j=1}^{d} (c_i)_j v_j = \sum_{j=1}^{k} (z_i)_j v_j + \sum_{m=k+1}^{d} (\xi_i)_m v_m. \quad (3)$$

We have shown that the coordinates which yield an optimal lower dimensional encoding are an orthogonal projection of the form

$$(V^T V)^{-1} V^T x_i = V^T x_i \Rightarrow (c_i)_j = v_j^T x_i,$$

where we use the fact that $V$ has orthonormal columns to simplify to $V^T x_i$. So, we can write (3) as

$$x_i = \sum_{j=1}^{k} (v_j^T x_i) v_j + \sum_{m=k+1}^{d} (v_m^T x_i) v_m$$

Therefore, the optimal linear autoencoder will minimize the norm of

$$x_i - \sum_{j=1}^{k} (v_j^T x_i) v_j = \sum_{m=k+1}^{d} (v_m^T x_i) v_m, \quad (4)$$

or the difference between $x_i$ and its $k$-dimensional encoding. It follows that (2) may be reformulated as

$$\min_{\{v_{k+1}, \dots, v_d\}} \frac{1}{n} \sum_{i=1}^{n} \left\| \sum_{m=k+1}^{d} (v_m^T x_i) v_m \right\|_2^2.$$

Notice that since the $v_m$ are mutually orthonormal, we can write

$$\left\| \sum_{m=k+1}^{d} (v_m^T x_i) v_m \right\|_2^2 = \left( \sum_{m=k+1}^{d} v_m^T x_i \right)^2 = \sum_{m=k+1}^{d} (v_m^T x_i)^2,$$

So it follows that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \| x_i - W_D W_D^{\dagger} x_i \|_2^2 &= \sum_{i=1}^{n} \sum_{m=k+1}^{d} (v_m^T x_i)^2 \\ &= \sum_{m=k+1}^{d} v_m^T \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right) v_m \quad (5) \\ &= \mathrm{Tr}(\bar{V}^T (X X^T) \bar{V}) \\ &= \mathrm{Tr}(\bar{V} \bar{V}^T (X X^T)). \end{aligned}$$

$\bar{V} \in \mathbb{R}^{d \times (d-k)}$ and $X X^T$ is the covariance matrix of our data (assuming it is mean-centered). Thus, minimizing (2) is equivalent to minimizing reconstruction error when orthogonally projecting the column space of the covariance matrix onto a $k$-dimensional subspace, which via (4), is the orthogonal complement to $\mathrm{Col}(\bar{V})$.

Write $\tilde{V} \in \mathbb{R}^{d \times k}$ as the matrix whose columns span this $k$-dimensional subspace

From (5) and the fact that $v_1, \dots, v_d$ are orthonormal, this minimization decouples into a sum of Rayleigh-Quotients of the form

$$\sum_{m=k+1}^{d} v_m^T \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right) v_m = \sum_{m=k+1}^{d} v_m^T (XX^T) v_m$$

This forces $\mathrm{Col}(\tilde{V})$ to be the span of the eigenvectors of $XX^T$ corresponding to the smallest $d-k$ eigenvalues. Hence, $\mathrm{Col}(\tilde{V})$ must be the subspace spanned by the eigenvectors of $XX^T$ corresponding to the largest $k$ eigenvalues.

This is precisely the desired result.

We used a two-step minimization. In one step, we showed that the linear autoencoder must be an orthogonal projection. In the other, we were able to determine which basis upon which an orthogonal projection yields an optimal solution to the autoencoder minimization.

Finally, note that we need to use a two-step minimization since, for example, if we write $A = W_D W_E \in \mathbb{R}^{d \times d}$, then solving

$$\min_{A} \frac{1}{n} \sum_{i=1}^{n} \| x_i - A x_i \|_2^2$$

will yield $I^{d \times d}$ as the trivial soln, and we would not be able to learn anything about the encoding and decoding maps. □

(b) Provide one explicit soln for $W_D$ and $W_E$. Argue why this is not unique.

Soln: Using the notation from above, pick $W_D = \tilde{V}$ and $W_E = \tilde{V}^T$. This is not unique since, for any matrix $Q \in \mathbb{R}^{k \times k}$ with orthonormal columns, we have

$$(\tilde{V}Q)(\tilde{V}Q)^T = \tilde{V}QQ^T\tilde{V}^T = \tilde{V}\tilde{V}^T = W_D W_E,$$

implying that the mapping $(\tilde{V}Q)(\tilde{V}Q)^T$ is also an optimal linear autoencoder. □

(c) What is the soln if we additionally require $W_D$ and $W_E^T$ to have orthonormal columns.

Soln: My simplifying assumption of orthonormality in (a) already hinted at this. In this case, we have

$$W_D W_E = W_D W_D^{\dagger} = W_D W_D^T = \tilde{V}\tilde{V}^T,$$

where $\tilde{V} \in \mathbb{R}^{d \times k}$ is the matrix whose columns have equivalent span to the top $d-k$ eigenvectors of $XX^T$ and are orthonormal. So this is a PCA projection/embedding onto the subspace spanned by the principal components.

(a) Explain how parts (a) - (c) relate to PCA.

Soln: Using the notation from above, we can write

$$\min_{\{v_{k+1},\dots,v_d\}} \frac{1}{n}\sum_{i=1}^{n} \|x_i - \tilde{V}\tilde{V}^T x_i\|_2^2 = \min_{\{v_{k+1},\dots,v_d\}} \frac{1}{n}\sum_{i=1}^{n} \|(I - \tilde{V}\tilde{V}^T)x_i\|_2^2.$$

That is, we want the best linear, rank-k approximation to the identity map. This is exactly what PCA does. In other words, PCA is an optimal linear autoencoder. □

② Let $x_1, \dots, x_n \in \mathbb{R}^d$ be unit vectors representing $n$ distinct lines in $\mathbb{R}^d$. Assume the lines are equiangular, i.e. $\exists \alpha \in (0,1)$ s.t. $|\langle x_i, x_j \rangle| = \alpha$ for all $i \neq j$.

(a) For each $i$, define the rank-one projection $P_i = x_i x_i^T \in \mathrm{Sym}_d$ where $\mathrm{Sym}_d$ denotes the space of real symmetric matrices. Show that

$$\langle P_i, P_j \rangle_F = \begin{cases} 1, & i = j \\ \alpha^2, & i \neq j \end{cases}$$

s.t. $\langle A, B \rangle_F$ is the Frobenius inner inner product $\langle A, B \rangle_F = \mathrm{Tr}(A^T B)$.

Pf: First, note that column $[P_i]_{*k}$ of $P_i$ is of the form $[P_i]_{*k} = (x_i)_k \vec{x}_i$. Therefore, we see that

$$\langle P_i, P_j \rangle_F = \sum_{k=1}^d \langle [P_i]_{*k}, [P_j]_{*k} \rangle = \sum_{k=1}^d (x_i)_k (x_j)_k \langle \vec{x}_i, \vec{x}_j \rangle \quad (*)$$

where $\langle \cdot, \cdot \rangle$ is the inner product inducing the Euclidean vector norm. Assuming $i \neq j$, then $(*)$ reduces to

$$\sum_{k=1}^d (x_i)_k (x_j)_k \langle \vec{x}_i, \vec{x}_j \rangle = \langle \vec{x}_i, \vec{x}_j \rangle \sum_{k=1}^d (x_i)_k (x_j)_k = \langle \vec{x}_i, \vec{x}_j \rangle^2 = \alpha^2.$$

On the other hand, if $i = j$, then

$$\begin{aligned} \sum_{k=1}^d (x_i)_k (x_j)_k \langle \vec{x}_i, \vec{x}_j \rangle &\overset{(WLOG)}{=} \sum_{k=1}^d (x_i)_k^2 \langle \vec{x}_i, \vec{x}_i \rangle \\ &= \|x_i\|_2^2 \sum_{k=1}^d (x_i)_k^2 \\ &= \|x_i\|_2^2 \cdot \|x_i\|_2^2 \\ &= 1 \end{aligned}$$

since $\|x_i\| = 1 \ \forall i$. □

(b) Let $G$ be the Gram matrix of $\{P_1, \dots, P_n\}$ wrt the Frobenius inner product. Write $G$ explicitly and compute its eigenvalues. Deduce that $\{P_1, \dots, P_n\}$ is a LI set in $\mathrm{Sym}_d$.

Pf: We can express $G$ entrywise using the result in (a). That is, for $1 \leq i, j \leq n$ we have

$$G_{ij} = \langle P_i, P_j \rangle_F = \begin{cases} 1, & i = j \\ \alpha^2, & i \neq j \end{cases}.$$

This implies that we can write

$$G = \alpha^2 \vec{1}\vec{1}^T - (\alpha^2 - 1) I.$$

Therefore, if $Gv = \lambda v$ for some $v \in \mathbb{R}^n$, then $Gv = \alpha^2 \vec{1}\vec{1}^T v - (\alpha^2 - 1) v = \lambda v$.

First, notice that if $V = \vec{1}$ we obtain

$$G\vec{1} = \alpha^2 \vec{1}\vec{1}^T\vec{1} - (\alpha^2 - 1)\vec{1} = (\alpha^2 n - (\alpha^2 - 1))\vec{1}.$$

So $\lambda = 1 + (n-1)\alpha^2$ is an eigenvalue with a multiplicity 1. Additionally, if $V \perp \vec{1}$, then

$$GV = \alpha^2 \vec{1}\vec{1}^T V - (\alpha^2 - 1)V = (\alpha^2 \cdot 0)\vec{1} - (\alpha^2 - 1)V = -(\alpha^2 - 1)V$$

Let $W = \{V \in \mathbb{R}^n : \vec{1}^T V = 0\}$. Then $W$ is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. For example, note that

$$\left\{ \begin{bmatrix} 1 \\ -e_k \end{bmatrix} \right\}_{k=1}^{n-1}$$

where $e_k \in \mathbb{R}^{n-1}$ is a basis for $W$. Thus, $(\alpha^2 - 1)$ is an eigenvalue of $G$ with multiplicity $n-1$.

Thus, $G$ has no trivial eigenvalues, implying $G$ is invertible and therefore has linearly independent columns and rows.

To conclude, we WTS that the symmetric linear operator $P = \sum_{i=1}^n c_i P_i$ satisfies $PV = \vec{0}$ iff $\forall i : c_i = 0$, i.e. that the set $\{P_1, \dots, P_n\}$ is LI.

The direction $\forall i : c_i = 0 \implies PV = \vec{0}$ is immediate. → for contradiction

Now let $P = \sum_{i=1}^n c_i P_i = 0$ with some index $i$ s.t. $c_i \neq 0$. Since $\mathrm{Sym}_d$ is an IPS under $\langle A, B \rangle_F$, then we will use nondegeneracy and bilinearity to see that

$$0 = \langle P, P \rangle_F = \sum_{i=1}^n c_i \langle P_i, P \rangle_F = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \langle P_i, P_j \rangle_F.$$

Now we can relate this directly to the Gram matrix $G$. That is,

$$0 = \sum_{i=1}^n \sum_{j=1}^n c_i c_j G_{ij} = \sum_{i=1}^n c_i \vec{c}_J^T G_{i*} = \vec{c}_J^T \left( \sum_{i=1}^n c_i G_{i*} \right),$$

where $c_J \in \mathbb{R}^n$ is a constant vector collecting the scalars $c_j$ and $G_{i*}$ are row vectors of $G$. Since we started with $\langle P, P \rangle_F$, then $\forall 1 \leq i \leq n : (c_J)_i = c_i$. Together w/ invertibility of $G$, then

$$0 = \vec{c}_J^T \left( \sum_{i=1}^n c_i G_{i*} \right) = \sum_{i=1}^n c_i^2 G_{i*} \neq 0 \quad \text{↯}$$

$\therefore \{P_1, \dots, P_n\}$ are LI. □

(c) Use the fact that $\dim(\mathrm{Sym}_d) = \frac{d(d+1)}{2}$ to conclude that $n \geq \frac{d(d+1)}{2}$

Soln: This is immediate from (b). In words, we cannot choose more LI $P_i \in \mathrm{Sym}_d$ than $\dim(\mathrm{Sym}_d)$. This is the Gerzon bound.

(d) Suppose that the $\{P_i\}$ span $\text{Sym}_d$. Explain why it is impossible to add an additional rank-one projection $yy^T$ that is equiangular with all $P_i$.

Soln: We showed in (b) that the set of rank-one projectors $\{P_i\}$ onto a set of $n$ equiangular lines $\{x_i\} \subset \mathbb{R}^d$ form a LI set of linear, symmetric operators on $\mathbb{R}^d$, of which there are $\frac{1}{2}d(d-1)$ unique operators. In the case that the $\{P_i\}$ span $\text{Sym}_d$, then

$$n = \frac{d(d-1)}{2}.$$

Thus, if $\exists y \in \mathbb{R}^d$ equiangular with $\{x_i\}$, then the bound is violated, as our argument in (b) would imply that

$$\{y\} \cup \{x_i\} \text{ is a LI set} \Rightarrow n+1 = \frac{d(d-1)}{2}$$

In other words, the Gerzon bound provides an upper bound on the # of equiangular lines in $\mathbb{R}^d$, and this bound is precisely $\dim(\text{Sym}_d)$. □

(e) Argue that an equiangular set of unit vectors achieving equality in the Welch bound must satisfy

$$n \leq \frac{d(d+1)}{2}.$$

Soln: Repeat the argument in (b) and reasoning in (d).

Also, note that we have equality in the Welch bound when $n = \frac{d(d+1)}{2}$.

See Thm 5.10 of Foucart and Rauhut. □

(4) Let $A = [a_1, \dots, a_m] \in \mathbb{R}^{d \times m}$ have unit norm columns, i.e. $\|a_j\|_2 = 1, \forall j$. For an index set $S \subseteq [m]$ with $|S| = s$, let $A_S \in \mathbb{R}^{d \times s}$ denote the submatrix containing the columns of $A$ indexed by $S$.

**Def:** Say that a matrix $A \in \mathbb{R}^{d \times m}$ has the restricted isometry property (RIP) of order $s$ when $\exists \delta_s \in (0,1)$ s.t.

$$(1-\delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_s)\|x\|_2^2 \quad (1)$$

$\forall x \in \mathbb{R}^m$ that are $s$-sparse. Call the minimal $\delta_s$ for which this holds the isometry constant.

(a) Fix any set $S \subset [m]$ with $|S| = s$. Show that RIP of order $s$ implies that for all $u \in \mathbb{R}^s$

$$(1-\delta_s)\|u\|_2^2 \leq \|A_S u\|_2^2 \leq (1+\delta_s)\|u\|_2^2. \quad (2)$$

**Pf:** By restricting $A$ to $A_S$ and $x$ to $x_S$, the multiplication $A_S x_S$ can be seen as mapping an $s$ dimensional subspace of $\mathbb{R}^m$ to $\mathbb{R}^d$. That is, an $s$-sparse vector must lie in an $s$-dimensional subspace of $\mathbb{R}^m$. Conversely, any $u \in \mathbb{R}^s$ maps uniquely to an $s$-sparse vector $x \in \mathbb{R}^m$, up to choosing $S \subset [m]$. Since we are assuming fixed $S$, then there is a bijective correspondence.

Let $V_S \subset \mathbb{R}^m$ be the subspace spanned by $s$-sparse $x \in \mathbb{R}^m$. We know that (1) holds $\forall x \in V_S$. Further, we have argued that for fixed $S$, $x \in V_S$ are in bijection with $u \in \mathbb{R}^s$. Explicitly, for $i \in [s]$ we have $(x_S)_i = u_i$.

Thus, $x \in V_S \Rightarrow Ax = A_S u$ for some $u \in \mathbb{R}^s$. So (1) $\Rightarrow \|A_S u\|_2^2 \leq (1+\delta_s)\|x\|_2^2$.

But $x \in V_S \Rightarrow (x_S)_i = u_i \Rightarrow \|x\|_2^2 = \sum_{k=1}^m x_k^2 = \sum_{i \in [s]} u_i^2 + \underbrace{\sum_{k \notin S} x_k^2}_{=0} = \|u\|_2^2$.

$\therefore \|A_S u\|_2^2 \leq (1+\delta_s)\|u\|_2^2$ as required, and the LHS inequality follows similarly. □

(b) Deduce that the Gram matrix $G_S = A_S^T A_S$ satisfies the bound

$$(1-\delta_s)I_s \leq G_S \leq (1+\delta_s)I_s \quad (3)$$

for $I_s$ the $s \times s$ identity matrix. Conclude that the eigenvalues of $G_S$ lie in the interval $[1-\delta_s, 1+\delta_s]$.

**Pf:** To show that (3) holds, it is enough to show that

$$0 \leq G_S - (1-\delta_s)I_s,$$

i.e. that the symmetric matrix $G_S - (1-\delta_s)I_s \in \mathbb{R}^{s \times s}$ is PSD. In showing this, it would follow that

$$(1-\delta_s)I_s \preceq G_s,$$

and an identical argument for the symmetric matrix $(1+\delta_s)I_s - G_s \in \mathbb{R}^{s\times s}$ would establish the bound

$$G_s \preceq (1+\delta_s)I_s$$

on the RHS. We have

$$(1-\delta_s)\|x\|_2^2 \le \|Ax\|_2^2$$

$$u^T(1-\delta_s)u \le u^T A_s^T A_s u \quad [\text{from part (a)}]$$

$$0 \le u^T(G_s - (1-\delta_s)I_s)u$$

$$0 \preceq G_s - (1-\delta_s)I_s$$

$\forall u \in \mathbb{R}^s$, which proves (3). This bound together with (2) implies that

$$1-\delta_s \le \frac{u^T G_s u}{u^T u} \le 1+\delta_s,$$

$\forall u \in \mathbb{R}^s \setminus \{\vec{0}\}$. So, the eigenvalue result follows from Courant-Fischer.

(c) Using (b) explain why $A_s^T A_s$ is invertible for $\delta_s < 1$. Moreover, show that

$$\|(A_s^T A_s)^{-1}\|_2 \le \frac{1}{1-\delta_s}$$

Pf: If $\delta_s < 1$, then the bound established in (b) implies that

$$|\lambda_k(A_s^T A_s)| > 0$$

for any $k^{\text{th}}$ eigenvalue. Thus, $A_s^T A_s$ must be invertible since $\lambda_k(A_s^T A_s) \neq 0$ $\forall k$.

Let $M \in \mathbb{R}^{s\times s}$ be invertible. We will use the following two facts:

- $Mv = \lambda v \iff \frac{1}{\lambda} v = M^{-1} v$ i.e. $M^{-1}$ has identical eigenvectors with reciprocal eigenvalues (F1)
- $\|M\|_2^2 = \sigma_{max}(M) = \lambda_{max}(M^T M)$ (F2)

For $\delta_s \in (0,1)$ we obtain $\lambda_{max}((A_s^T A_s)^{-1}) = 1/\lambda_{min}(A_s^T A_s) = \frac{1}{1-\delta_s}$ via F1 and the bound in (b).

Using F2, then $\|(A_s^T A_s)^{-1}\|_2^2 = \frac{1}{1-\delta_s}$.

Taking square roots, we see that $\left(\frac{1}{1-\delta_s}\right)^{1/2} = 1/(1-\delta_s)^{1/2} \le \frac{1}{1-\delta_s}$ as $(1-\delta_s) \in (0,1)$.

So we obtain $\|(A_S^T A_S)^{-1}\|_2 = \left(\frac{1}{1-\delta_S}\right)^{1/2} \leq \frac{1}{1-\delta_S}$ as required. □

(a) Assume $A$ satisfies RIP of order $s \geq 2$ with constant $\delta_s$. Show that for any $i \neq j$ $|\langle a_i, a_j \rangle| \leq \delta_s$ and conclude that $\mu(A) \leq \delta_s$.

Pf: Note that we assume $s \geq 2$ as $\delta_s = 0$ for $s = 1$. Also, $\langle a_i, a_j \rangle = [A^T A]_{ij}$. Assume (WLOG) fixed $S \subset [m]$ for $|S| = s$. For each pair $i, j \in [m] \cap S$, we have some $k, l \in [s]$ s.t.

$$[A_S^T A_S]_{kl} = \langle a_i, a_j \rangle.$$

Using (3) we established that for any $k \in [s]$ we have

$$1 - \delta_s \leq \lambda_k(A_S^T A_S) \leq 1 + \delta_s \iff \lambda_k(A_S^T A_S) \in D(1, \delta_s) \quad (4)$$

Regardless of our choice of $S$, we have $\operatorname{diag}(A_S^T A_S) = \vec{1}$. Fixing $i \in [m] \cap S$, write

$$R_i = \sum_{\substack{j \in [m] \cap S \\ i \neq j}} |\langle a_i, a_j \rangle|.$$

Applying Gershgorin's Circle Thm, we see that $\lambda_k(A_S^T A_S) \in D(1, R_i)$ for each $i$. Taking this together with (4)

$$R_i \leq \delta_s \Rightarrow |\langle a_i, a_j \rangle| \leq \delta_s \ \forall\, i, j \in [m] \cap S.$$

As this analysis holds for any choice of $S$, the result follows. □

⑥ Given $x_1, \dots, x_k \in \mathbb{R}^d$ and a similarity/weight matrix $W$, the Laplacian is $L = D - W$ where $D \in \mathbb{R}^{k \times k}$ is the diagonal degree matrix. Assume that $D_{ii} \neq 0$ for $1 \leq i \leq k$. Consider

$$\min_{Y \in \mathbb{R}^{k \times m}} \operatorname{tr}(Y^T L Y)$$

where each row of $Y$ is an $m$ dimensional embedding of some $x \in \mathbb{R}^d$.

(a) What is the importance of the constraint $Y^T Y = I$?

Soln: Denote $\{c_i\}_{i=1}^m$ as the rows of $Y^T$. The constraint $Y^T Y = I$ implies that $\langle c_i, c_j \rangle = \delta_{ij}$. In words, that the columns of $Y$ are mutually orthonormal. Note that we can write

$$\operatorname{tr}(Y^T L Y) = \sum_{i=1}^m c_i^T L c_i = \sum_{i=1}^m \langle c_i, L c_i \rangle \quad (1)$$

and recall that $\lambda = 0$ is always an eigenvalue of $L$ with corresponding eigenvector $\vec{1}$. So, without the constraint $Y^T Y$, we can choose $Y = \vec{1}\vec{1}^T$ so that $c_i = \vec{1}$ for each $i$.

Since $\langle c_i, L c_i \rangle \geq 0$ for each $i$, then $\min_{Y \in \mathbb{R}^{k \times m}} \operatorname{tr}(Y^T L Y) = \vec{1}\vec{1}^T \in \mathbb{R}^{k \times a}$, as

$$Y = \vec{1}\vec{1}^T \implies \operatorname{tr}(Y^T L Y) = \sum_{i=1}^m \vec{1}^T L \vec{1} = \sum_{i=1}^m \vec{1}^T \vec{0} = 0 \quad \square$$

* each embedding is identical!

(b) Derive the Lagrangian for the constrained optimization

$$\min_{Y \in \mathbb{R}^{k \times m}} \operatorname{tr}(Y^T L Y) \text{ s.t. } Y^T Y = I \quad (2)$$

Soln: Define $F, G : \mathbb{R}^{m \times m} \to \mathbb{R}$ as $F(Y) = \operatorname{tr}(Y^T L Y)$ and $G(Y) = \operatorname{tr}(Y^T Y - I)$. Now let $Y_*$ be an optimal solution to

$$\min_{Y \in \mathbb{R}^{k \times m}} F(Y) \text{ s.t. } G(Y) = 0,$$

which is equivalent to (2). Since $F, G$ both decouple into a sum of inner products, then they are both $C^1$ functions. So, via the Lagrange Multiplier Thm, $\exists \Lambda \in \mathbb{R}^{m \times m}$, s.t.

$$\nabla F(Y_*) = \Lambda \nabla G(Y_*).$$

Decoupling into a sum over subproblems, then the minimization is

$$\min_{c_1, \dots, c_m} \sum_{i=1}^m c_i^T L c_i \text{ s.t. } \forall i : c_i^T c_i - 1 = 0, \quad (3)$$

and each subproblem has a Lagrangian $2Lc_i = 2\lambda_i c_i \iff Lc_i = \lambda_i c_i$ $\square$

(c) Show that any optimal solution $Y_*$ consists of eigenvectors of $L$. Which should we choose to minimize the objective?

Soln: Since the Lagrangian is of the form $Lc_i = \lambda_i c_i$, then this is in fact an eigenvalue problem for the matrix $L$. Furthermore, since $c_i^T c_i = 1$ for each $i$, then we are solving

$$\min_{c_1,\dots,c_m} \sum_{i=1}^m c_i^T L c_i = \min_{c_1,\dots,c_m} \sum_{i=1}^m \frac{c_i^T L c_i}{c_i^T c_i},$$

or minimizing a sum of Rayleigh Quotients.

Via the Rayleigh-Ritz Thm, the eigenvectors $c_i^*$ which satisfy (3) must correspond to the first $m$ nonzero eigenvalues of $L$ with the smallest magnitude. □

(d) Consider an alternative constraint: $Y^T D Y = I$. Derive the optimality conditions for

$$\min_{Y \in \mathbb{R}^{k\times m}} \operatorname{tr}(Y^T L Y) \text{ s.t. } Y^T D Y = I. \quad (4)$$

Show that that the columns of an optimal solution $Y_*$ satisfy the generalized eigenvalue problem $Ly = \lambda Dy$.

Note: I have never worked with KKT / optimality conditions. I used the following text as a reference:

Ghojogh et. al. KKT Conditions, First-Order and Second-Order Optimization, and Distributed optimization. University of Waterloo.

Pf: It will be enough to determine the stationarity and primal feasibility conditions.

Assume a fixed vector $c \in \{c_i\}_{i=1}^m$; i.e. some row of $Y^T$. Since (4) decouples into $m$ identical subproblems of the form

$$\min_{c \in \mathbb{R}^k} c^T L c \text{ s.t. } c^T D c - 1 = 0,$$

we will determine the optimality conditions for each.

Assuming that $c_* \in \mathbb{R}^k$ is an optimal solution, then the stationarity condition is

$$\nabla_c (c_*^T L c_*) + \lambda \nabla_c (c_*^T L c_*) = 0 \iff 2Lc_* = 2\lambda D c_* \iff Lc_* = \lambda D c_* \quad (5)$$

for some $\lambda \in \mathbb{R}$. So, the Lagrangian is exactly the generalized eigenvalue problem.

Note that the primal feasibility condition is $c_*^T D c_* - 1 = 0$.

$\therefore$ Any column $c_* \in \mathbb{R}^k$ of the optimal solution $Y_* \in \mathbb{R}^{k\times m}$ must satisfy $Lc_* = \lambda D c_*$ □

(e) Explain which generalized eigenvectors should be chosen to form $Y_*$. How does this choice depend on the embedding dimension $m$?

Soln: Rewrite (5) as $LY_* = DY_*\Lambda$ where $Y_* \in \mathbb{R}^{k\times m}$ collects the generalized eigenvectors $c_*$ along its columns and $\Lambda \in \mathbb{R}^{m\times m}$ is a diagonal matrix containing the eigenvalues. Since $D_{ii} \neq 0 \; \forall i$, then we can write this as a standard eigenvalue problem

$$D^{-1}LY_* = Y_*\Lambda \iff D^{-1}Lc_* = \lambda c_*.$$

So, we are solving

$$\min_{c_1,\dots,c_m} \sum_{i=1}^{m} \frac{c_i^T D^{-1} L c_i}{c_i^T c_i}.$$

Applying Rayleigh-Ritz again, the eigenvectors $c_i^*$ which satisfy this must correspond to the first $m$ nonzero eigenvalues of $D^{-1}L$ with the smallest magnitude.

Another way to interpret this constraint is the orthonormality constraint $\langle c_i, Dc_j \rangle = \delta_{ij}$.

However, we need $m$ generalized eigenvectors to form the columns of $Y_* \in \mathbb{R}^{k\times m}$ since we are using them to construct $m$-dimensional embeddings. It is natural to assume that $m \geq d$, or that this is a dimensionality reduction. On the other hand, given a fixed choice for $m$, we require $k \geq m$ to construct the embedding since $D^{-1}L$ and $L$ can have no more than $k$ (generalized) eigenvectors. □

(f) Exploration: Do some research and briefly discuss how this finite dimensional problem relates to eigenfunction problems of the Laplace-Beltrami operator in the continuous setting.

Soln: Most importantly, the graph Laplacian can be seen as a discretization of the Laplace equation

$$\Delta f = \sum \frac{\partial^2 f}{\partial x_i^2} \quad \text{for } f: \mathbb{R}^n \to \mathbb{R}.$$

The simplest example is the path graph, where each row of its graph Laplacian is the centered difference quotient

$$\text{Row}_i\left(\tfrac{1}{2}Lf\right) = \frac{-f_{i-1} + 2f_i - f_{i+1}}{2} = f''(x_i)$$

of a function $f: \mathbb{R} \to \mathbb{R}$. In this case, the path graph has vertices $X = \{x_i\}$ with a uniform mesh of an interval $[a,b]$ s.t. $h = \sqrt{2}$. We can use this to solve certain ODEs and PDEs, or compute numerical derivatives.

Laplace-Beltrami generalizes the homogeneous PDE $\Delta f = 0$ to a function space where each function is defined on a manifold. Just as the eigenvectors of the graph Laplacian were related to the geometry of the data represented by the graph, the eigenfunctions of Laplace-Beltrami, or solutions to the Helmholtz operator $\Delta f = \lambda f$, are well-adapted to the geometry of the manifold. □

(g) Bonus: Suppose we add the constraint $Y^T D\vec{1} = 0$ to the problem in D. Explain how this affects the optimal solution.

⑦ Find all the eigenvalues and characterize the corresponding eigenvectors for the Laplacian of the complete graph.

Soln: Let $G = (V, E)$ be a complete graph with $|V| = n$. Then the graph Laplacian $L = D - A$ has the form:

- $D \in \mathbb{Z}_+^{n \times n}$ is a diagonal degree matrix s.t. $\text{diag}(D) = n\vec{1}$
- $A \in \mathbb{Z}_2^{n \times n}$ is a complete adjacency matrix, i.e. $A_{ij} = 1$ $\forall$ pairs $i \neq j$

We have shown in lecture that $0 \in \mathbb{R}$ and $\vec{1} \in \mathbb{R}^n$ is an eigenpair for the Laplacian of any graph with $n$ vertices. Explicitly,

$$(L\vec{1})_i = (D\vec{1})_i - (A\vec{1})_i = \deg(v_i) - \sum_{i:(i,j)\in E} 1 = \deg(v_i) - \deg(v_i) = 0.$$

As this holds entrywise, then $L\vec{1} = \vec{0} \Rightarrow L\vec{1} = 0\vec{1}$.

Let $\vec{x} \in \mathbb{R}^n$. Note that when the graph is complete we have

$$L = D - A = (n-1)\cdot I_n - (\vec{1}\vec{1}^T - I_n) = n \cdot I_n - \vec{1}\vec{1}^T.$$

The spectrum of $n \cdot I_n$ is summarized as follows:

- $n \cdot I_n$ has an eigenvalue $n$ with multiplicity $n$.
- $\forall x \in \mathbb{R}^n$: $(n \cdot I_n)\vec{x} = n \cdot (I_n \vec{x}) = n \cdot \vec{x} \Rightarrow \vec{x}$ is an eigenvector of $n \cdot I_n$.
- * We will choose the following basis for this $n$-dimensional eigenspace:

$$\vec{1} \cup \{v \in \mathbb{R}^n : \vec{1}^T v = 0\} = \vec{1} \cup W$$

The motivation for this choice will soon be clear.

The spectrum of $\vec{1}\vec{1}^T$ is summarized as follows:

- For $x = \vec{1}$ we have $\vec{1}\vec{1}^T x = n \cdot \vec{1}$; so $n$ is an eigenvalue of $\vec{1}\vec{1}^T$ with eigenvector $\vec{1}$ and multiplicity 1.
- For $\vec{x} \perp \vec{1}$ we have $\vec{1}\vec{1}^T x = 0 \cdot \vec{1} = 0$; so 0 is an eigenvalue of $\vec{1}\vec{1}^T$ with multiplicity $(n-1)$. To be explicit, $W$ is the eigenspace of $\vec{1}\vec{1}^T$ associated with eigenvalue 0.

Now, recall that for $A\vec{v} = \lambda\vec{v}$ and $B\vec{z} = \gamma\vec{z}$ that if $A+B$ is compatible and $A, B$ share an eigenvector $\vec{v}$ (i.e. $\vec{v} = \vec{z}$), then $(A+B)\vec{v} = (\lambda + \gamma)\vec{v}$. (1)

Finally, we can classify the spectrum of $L$. By our choice of $W$, we can use (*) to see that:

- $L\vec{1} = n \cdot I_n \vec{1} - \vec{1}\vec{1}^T\vec{1} = (n-n)\vec{1} = 0 \cdot \vec{1}$; so $0$ is an eigenvalue with eigenvector $\vec{1}$ as expected.

- If $\vec{x} \in W$ then $n \cdot I_n \vec{x} - \vec{1}\vec{1}^T\vec{x} = (n-0)\vec{x} = n \cdot \vec{x}$; so $n$ is an eigenvalue with multiplicity $(n-1)$, and the corresponding eigenvectors are perpendicular to $\vec{1}$.

* $L$ has one connected component as expected, and $(n-1)$ ideal cuts. □

(8) The quadratic form of the Laplacian is $x^T L x = \frac{1}{2}\sum_{(i,j)\in E} A_{ij}(x_i - x_j)^2$.

Consider an eigenvector $v$ corresponding to the the second smallest eigenvalue, $\lambda_2$. Given that $v \perp \vec{1}$, use the interpretation of the quadratic form to argue which would have a larger $\lambda_2$: a path graph ($P_n$) or a complete graph ($C_n$)?

Soln: Let $L_P, L_C : \mathbb{R}^n \to \mathbb{R}$ be quadratic forms such that

$$L_P(x) = x^T[L_P]x \quad \text{and} \quad L_C(x) = x^T[L_C]x$$

where $[L_P], [L_C] \in \mathbb{Z}^{n\times n}$ are the graph Laplacian matrices of $P_n$ and $C_n$. To each we can associate a fiedler vector $v_P, v_C \in \mathbb{R}^n$ s.t.

$$v_P = \min_{v\in\mathbb{R}^n} \frac{1}{n} L_P(v) \quad \text{and} \quad v_C = \min_{v\in\mathbb{R}^n} \frac{1}{n} L_C(v)$$

where each objective is constrained by (1) $v \perp 1$ and (2) $v^T v = n$.

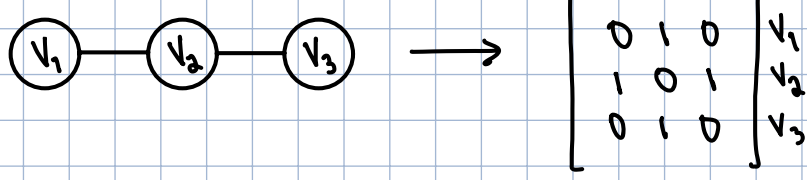
Constraint (2) means that each objective is a Rayleigh quotient, and (1) together with Courant-Fischer implies that $v_P$ and $v_C$ are eigenvectors which correspond to the second smallest eigenvalues of $[L_P]$ and $[L_C]$, respectively.

Denote the respective Fiedler values as $\lambda_P$ and $\lambda_C$. By construction, we can compare $\lambda_P$ and $\lambda_C$ by computing $L_P(v_P)$ and $L_C(v_C)$.

Using our work on Q7, we see that $\lambda_C = L_C(v_C) = n-1$ for each $n$.

On the other hand, observe that the adjacency matrix of the path graph, $A_P \in \mathbb{Z}_2^{n\times n}$, is a banded matrix with empty diagonal, assuming that $A_P$ is indexed in an ascending path order.

Aside: By ascending path order, I mean something like this:




End Aside.

In this case, we have $[A_P]_{ij} = \begin{cases} 1, & i-j = \pm 1 \\ 0, & \text{else} \end{cases}$.

Via symmetry of the graph Laplacian, we can write

$$L_P(v_P) = \frac{1}{2}\sum_{(i,j)\in E} [A_P]_{ij}\left(v_{P(i)} - v_{P(j)}\right)^2 = \sum_{\substack{(i,j)\in E \\ i>j}} [A_P]_{ij}\left(v_{P(i)} - v_{P(j)}\right)^2,$$

thus restricting to the case $i-j=1$. We can rewrite this as

$$\sum_{\substack{(i,j)\in E \\ i>j}} [A_P]_{ij}\left(v_{p(i)}-v_{p(j)}\right)^2 = \sum_{i=2}^{n}\left(v_{p(i)}-v_{p(i-1)}\right)^2. \quad (2)$$

Having written $\lambda_P$ in this form, we see that pairwise differences $v_{p(i)}-v_{p(i-1)}$ are minimized. Since $\lambda_P$ is assumed to be the minimal (nontrivial) solution. So, using this and by appealing to symmetry,* we see that $v_P$ also minimizes

$$\frac{1}{2}\left(\sum_{i=2}^{n} v_{p(i)}-v_{p(i-1)} + \sum_{i=1}^{n-1} v_{p(i+1)}-v_{p(i)}\right). \quad (3)$$

Now rearrange and use the fact that $v_P^T\vec{1}=0$:

$$= \frac{1}{2}\left[\left(\sum_{i=2}^{n} v_{p(i)}\right)-\left(\sum_{i=2}^{n} v_{p(i-1)}\right)+\left(\sum_{i=1}^{n-1} v_{p(i+1)}\right)-\left(\sum_{i=1}^{n-1} v_{p(i)}\right)\right]$$

$$= \frac{1}{2}\left[\left(v_P^T\vec{1}-v_{p(1)}\right)-\left(v_P^T\vec{1}-v_{p(n)}\right)+\left(v_P^T\vec{1}-v_{p(1)}\right)-\left(v_P^T\vec{1}-v_{p(n)}\right)\right]$$

$$= v_{p(n)}-v_{p(1)} \quad (4)$$

So, viewing this as a discretized PDE, we seek an eigenfunction where the absolute differences between the function evaluations along the boundary is minimized; i.e. $v_P$ is <u>close to</u> being periodic.

Note that we cannot have $v_{p(n)}-v_{p(1)}=0$, otherwise

$$(3)=0 \Rightarrow \frac{1}{2}\sum_{i=2}^{n} v_{p(i)}-v_{p(i-1)} = -\frac{1}{2}\sum_{i=1}^{n-1} v_{p(i+1)}-v_{p(i)}$$

* Via symmetry, each term of these sums is identical

$$\frac{1}{2}\sum_{i=2}^{n}\left(v_{p(i)}-v_{p(i-1)}\right)^2 = -\frac{1}{2}\sum_{i=1}^{n-1}\left(v_{p(i+1)}-v_{p(i)}\right)^2$$

$$\frac{1}{2}L_P(v_P) = -\frac{1}{2}L_P(v_P)$$

$$\Rightarrow L_P(v_P) = \lambda_P = 0$$

which contradicts the fact that $\lambda_P$ is the Fiedler value. On the other hand, we can observe that

- the condition that $v_P^T\vec{1}=0$ implies that $v_P$ has positive and negative entries
- Since $v_P$ minimizes (2) and (4), then pairwise differences are minimized and pairwise entries will often have the same sign. Interpretation: the value that $v_P$ takes varies slowly along the path, and $v_{p(n)}-v_{p(1)}$ is small

So, as $n$ grows, we expect $|v_{p(i)}-v_{p(i-1)}| \approx \frac{1}{n+1}$, up to a constant. Asymptotically,

$$\lambda_P \sim \sum_{i=1}^{n-1}\frac{1}{(n+1)^2} \sim \frac{n-1}{n^2+2n+1} \sim \frac{1}{n^2}. \quad \text{→ Scaling}$$

$\therefore P_n$ becomes less connected as $n$ grows, and $\lambda_P < \lambda_C$ in general. □

(10) Let $Y \in \mathbb{R}^{m \times n}$ be a data matrix with $n$ data points $y_i \in \mathbb{R}^m$, and consider the optimization problem

$$\min_{D \in \mathbb{R}^{m \times k},\, X \in \mathbb{R}^{k \times n}} F(D,X) = \min_{D \in \mathbb{R}^{m \times k},\, X \in \mathbb{R}^{k \times n}} \|Y - DX\|_F^2 + \lambda \|X\|_1$$

where $\lambda > 0$ and $m < k$ (overcomplete dictionary). In parts (b) - (f), additionally assume that $X \geq 0$.

(a) Show that $F(D,X)$ is convex in $X$ for fixed $D$ and convex in $D$ for fixed $X$.

Pf: First we will state a few lemmas.

Lemma 1: Let $f: \mathbb{R}^m \to \mathbb{R}$ be a convex function and $T: \mathbb{R}^n \to \mathbb{R}^m$ a linear mapping with $\text{im}\, T \subseteq \text{dom}\, f$. Then $(f \circ T): \mathbb{R}^n \to \mathbb{R}$

Lemma 2: Let $f_i: \mathbb{R}^n \to \mathbb{R}$ be convex functions with $\alpha_i \in \mathbb{R}$. Then $\sum_i \alpha_i f_i$ is a convex function.

Pf: Fix $D = A \in \mathbb{R}^{m \times k}$. We WTS that

$$F_A(X) = F(A,X) = \|Y - AX\|_F^2 + \lambda \|X\|_1$$

is convex. It is enough to show that $\|Y - AX\|_F^2$ and $\|X\|_1$ are convex, and the desired result would follow from Lemma 2.

First observe that for $\theta \in (0,1)$ we have

$$\|\theta X_1 + (1-\theta) X_2\|_1 \leq \|\theta X_1\|_1 + \|(1-\theta) X_2\|_1 = \theta \|X_1\|_1 + (1-\theta) \|X_2\|_1$$

via the $\Delta$-inequality. So $\|\cdot\|_1 : \mathbb{R}^{k \times n} \to \mathbb{R}$ is convex.

Now define $f_i: \mathbb{R}^m \to \mathbb{R}$ set $f_i(z) = \|y_i - z\|_2^2$ for $y_i$ the $i^{th}$ column of $Y$.

Consider $f(Ax_i)$ for $x_i$ the $i^{th}$ column of $X$. We can write

$$\|Y - AX\|_F^2 = \sum_{i=1}^n f_i(Ax_i). \quad (*)$$

For now, assume that $f_i(z)$ is convex. As $A$ is a linear mapping, then Lemma 1 implies that $f_i(Ax_i)$ is convex for each $i$. Applying Lemma 2, we see that $(*)$ is convex. Now we show that $f_i(z)$ is convex. Observe that

$$\nabla f_i(z) = 2(z - y_i) \Rightarrow \nabla^2 f_i(z) = 2 I^{m \times m},$$

which is a PSD matrix. So $F_A(X)$ is convex. Now define $F_H(D) = F(D,H)$ s.t. $X = H$ is fixed. In this case, we proceed by computing

$$\nabla^2 F_H(D) = \nabla^2 (\|Y - DH\|_F^2 + \lambda \|H\|_1) = \nabla^2 (\|Y - DH\|_F^2) = \nabla^2 (\|Y^T - H^T D^T\|_F^2).$$

which we obtain by linearity of $\nabla^2$ and the fact that $\lambda\|H\|_1$ is constant wrt $D$.

Now, by taking $\tilde{y}_j$ to be the $j^{th}$ column of $Y^T$ and $\tilde{d}_j$ the $j^{th}$ column of $D^T$ for $1 \leq j \leq m$, then we can define $g_j : \mathbb{R}^m \to \mathbb{R}$ s.t.

$$g_j(z) = \|\tilde{y}_j - z\|_2^2,$$

and we see that

$$\nabla^2 F_H(D) = \nabla^2\left(\sum_{j=1}^m g_j(H^T\tilde{d}_j)\right) = \sum_{j=1}^m \nabla^2(g_j(H^T\tilde{d}_j))$$

Using our argument from (*), we see that

$$\nabla^2 F_H(D) = \sum_{j=1}^m 2I^{n\times n} = 2mI^{n\times n},$$

which is PSD. So $F_H(D)$ is convex. □

(b) Explain why $X \geq 0$ preserves convexity of the $X$ subproblem.

Soln: We have shown that $F_A(X)$ is convex, and the constraint $X \geq 0$ simply restricts the space over which we are optimizing. In particular, for any $\theta \in (0,1)$ and $X_1, X_2 \in \mathbb{R}_+^{n\times n}$ we see that $\theta X_1 \geq 0$ and $(1-\theta)X_2 \geq 0$. Thus,

$$\theta X_1 + (1-\theta)X_2 \in \mathbb{R}_+^{n\times n} \quad \forall \theta \in (0,1).$$

∴ This is optimization of a convex function over a convex set, so the subproblem remains convex. □

(c) Explain why the problem is ill-posed if no constraint is imposed on $D$.

Soln: Without constraint on $D$, we can easily choose a dictionary $\tilde{D} \in \mathbb{R}^{m\times k}$ to represent $Y = \tilde{D}X$ exactly. This is due to the fact that $\tilde{D}$ is overcomplete, so its columns will span $\mathbb{R}^m$ with high probability. Given such $\tilde{D}$, pick $X$ s.t.

$$Y_{*j} = \tilde{D}X_{*j}.$$

To make this more explicit, suppose that $\{d_{i_d}\}_{d=1}^m$ is a LI subset of the columns of $\tilde{D}$ and that

$$Y_{*j} = \sum_{d=1}^m c_d d_{i_d}.$$

Then let $X_{ij} = c_d$ if $i = i_d$ for some $1 \leq d \leq m$. By constructing $X$ in this manner, then our problem reduces to sparse coding. For fixed $\tilde{D}$, we solve this by writing each $Y_{*j}$ wrt the basis along the columns of $\tilde{D}$ which produces the sparsest possible representation of $Y_{*j}$.

This constructive argument exemplifies that there are uncountably many dictionaries $\tilde{D}$ which exactly solve $Y = \tilde{D}X$ for some $X$, so the general problem is

intractable, as it decouples into uncountably many sparse coding problems.

On the other hand, by requiring $X \geq 0$ and $D$ without constraint, we can pick $\tilde{D}$ to be any dictionary for which each column $Y_{*i}$ appears as a column of $\tilde{D}$. In this case, construct $X$ s.t. $X_{ij} = 1$ if $\tilde{D}_{*j} = Y_{*i}$, and $X_{ij} = 0$ in any other case. Therefore, we have

$$\|Y - \tilde{D}X\|_F^2 + \lambda \|X\|_1 = 0 + \lambda = \lambda$$

since each $Y_{*i}$ can be written as a 1-sparse linear combination of columns of $\tilde{D}$ s.t. $Y = \tilde{D}X$. This solution is trivial, and there are uncountably many since $k - m$ columns of $\tilde{D}$ can be any vector in $\mathbb{R}^m$. □

(d) To address the issue in (c), suppose that we constrain $D$ to the set

$$\mathcal{D} = \{M \in \mathbb{R}^{m \times k} : \|M_{*j}\|_2 \leq 1 \text{ for all } j\},$$

i.e. the columns of $D$ have a Euclidean norm at most 1. Show that $F_\lambda(D)$ remains convex, and explain why this constraint is natural.

Pf: Let $M, N \in \mathcal{D}$ and $\theta \in (0,1)$. We will show that $\theta M + (1-\theta)N \in \mathcal{D} : \forall \theta \in (0,1)$. Similar to part (b) we are constraining the optimization of a convex function to a convex region.

Now, observe that for any $j^{th}$ column vector we have

$$\|\theta M_{*j} + (1-\theta)N_{*j}\|_2 \leq \theta \|M_{*j}\|_2 + (1-\theta)\|N_{*j}\|_2 \leq \theta + 1 - \theta = 1,$$

which implies $\theta M + (1-\theta)N \in \mathcal{D}$ as desired. □

This constraints is natural as it limits atoms to live in the unit ball in $\mathbb{R}^m$, which is a <u>bounded</u> and <u>convex</u> region.

(e) Consider the alternating minimization

$$X^{(t+1)} \in \underset{X \geq 0}{\operatorname{argmin}} \; F(D^{(t)}, X) \qquad (1)$$

$$D^{(t+1)} \in \underset{D \in \mathcal{D}}{\operatorname{argmin}} \; F(D, X^{(t+1)}) \qquad (2)$$

where each subproblem is solved exactly. Show that the objective values form a nonincreasing sequence.

Pf: Fix $t \in \mathbb{Z}_+$. We WTS that

$$F(D^{(t)}, X^{(t)}) \geq F(D^{(t+1)}, X^{(t+1)})$$

Since we have shown that the objectives in (1) and (2) are convex and

remain convex under their given constraints, then

$$F(D^{(t)}, X^{(t+1)}) \leq F(D^{(t)}, \tilde{X}) \quad \forall \tilde{X} \in \mathbb{R}^{n \times m} \text{ s.t. } \tilde{X} \geq 0, \quad (3)$$

and similarly,

$$F(D^{(t+1)}, X^{(t+1)}) \leq F(\tilde{D}, X^{(t+1)}) \text{ s.t. } \tilde{D} \in \mathcal{D} \quad (4)$$

With these facts we see that

$$F(D^{(t+1)}, X^{(t+1)}) \overset{(4)}{\leq} F(D^{(t)}, X^{(t+1)}) \overset{(3)}{\leq} F(D^{(t)}, X^{(t)}).$$

which is the desired result. □

(f) Show that the sequence $F(D^{(t)}, X^{(t)})$ converges, and explain why this does not imply convergence of the iterates $(D^{(t)}, X^{(t)})$.

Pf: Note that this is simply a real sequence which is nonincreasing and bounded below by zero. So the result follows from the monotone convergence theorem. □

Suppose that $F(D^{(t)}, X^{(t)}) \to l$. Then $\exists m \in \mathbb{Z}_+ : \forall n > m : F(D^{(n)}, X^{(n)}) = l$. Thus, any tail sequence $F(D^{(t+m)}, X^{(t+m)})$ is a constant sequence.

However, this is not enough to conclude that $D^{(t+m)} = D^{(t'+m)}$ and $X^{(t+m)} = X^{(t'+m)}$ for $t' \neq t$. This is due to the fact $F(D, X)$ is not jointly convex, and the subproblems $F_A(X)$ and $F_H(X)$ are not strictly convex. In other words, we cannot guarantee that a local minimum of $F(D, X)$ is a global minimum, or that it is unique.

## Question 9

The goal of this problem is to explore graph cuts and spectral clustering.

(a) Construct a connected graph of your choice with at least six nodes for which the minimum cut yields a trivial partition into two clusters. Compute the Fiedler vector (the eigenvector corresponding to the second-smallest eigenvalue of the graph Laplacian). Apply thresholding to the Fiedler vector and determine whether it correctly identifies the two clusters.

**Solution:** It does not correctly identify the clusters. See below.

In [62]:
```
# define the graph here
edges = [[0,1],[1,2],[2,3],[3,0],  [3,4],  [4,5],[5,6],[6,7],[7,4]]
G = nx.Graph()
G.add_edges_from(edges)

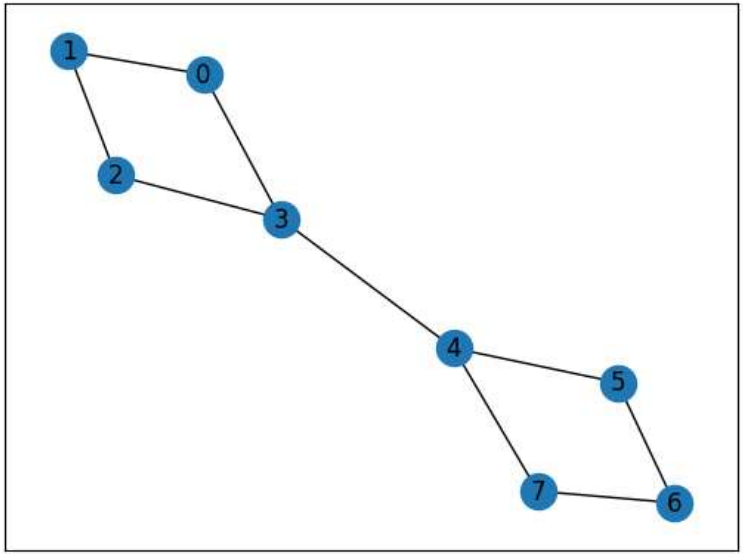
# plot the graph
nx.draw_networkx(G)
plt.show()

# construct its Laplacian
L = nx.laplacian_matrix(G).toarray()

# compute the fiedler vector
vals, vecs = sp.linalg.eigh(L)
fiedler_vec = vecs[1]

# thresholding - via an indicator function
cluster1 = [i for i, v in enumerate(fiedler_vec) if v <= 0]
cluster2 = [i for i, v in enumerate(fiedler_vec) if v > 0]

# plot the result
print(cluster1)
print(cluster2)
```




```
[0, 2, 5, 7]
[1, 3, 4, 6]
```

(b) Repeat the experiment in part (a) using a connected graph with at least nine nodes, where the minimum cut solution corresponding to three clusters is visually or structurally apparent. Compute the eigenvectors corresponding to the first two nonzero smallest eigenvalues of the graph Laplacian. Form an embedding using these eigenvectors and cluster the rows of the resulting eigenvector matrix to obtain three clusters. Does this procedure recover the underlying cluster structure?

**Solution:** I implemented this using the *sci-kit* k-means function for the clustering step. This procedure did recover the structure. Note that the labels of the printed result correspond exactly to the plotted graph. For example, nodes 3,4,5 and 9 are all given the label 2.

In [ ]:
```
# define the graph here
edges = [[0,1],[1,2],[2,0],  [3,4],[4,5],[5,3],  [6,7],[7,8],[8,6],  [0,9],[3,9],[6,9]]
G = nx.Graph()
G.add_edges_from(edges)
```

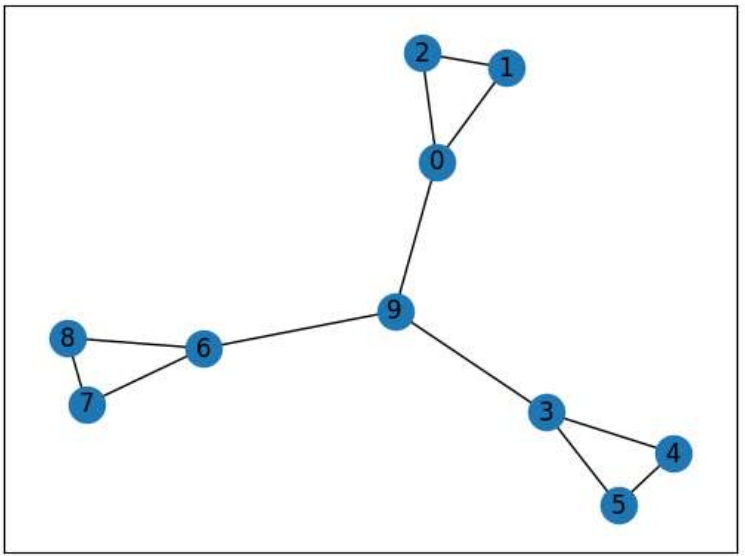
```
# plot the graph
nx.draw_networkx(G)
plt.show()

# construct its Laplacian
L = nx.laplacian_matrix(G).toarray()

# compute the two smallest (nonzero) eigenvalues
vals, vecs = sp.linalg.eigh(L)
V = vecs[:, 1:3]

# cluster using k-means
kmeans = KMeans(n_clusters=3, n_init=10)
labels = kmeans.fit_predict(V)

# return
print(labels)
```




```
[0 0 0 2 2 2 1 1 1 2]
```