

Solution to Problem 1.

(a)

i It is clear that $U_r \in \mathbb{R}^{m \times r}$, $\Sigma_r \in \mathbb{R}^{r \times r}$, and $V_r \in \mathbb{R}^{n \times r}$. [2]

ii $A^T A = V_r \Sigma_r^2 V_r^T$ and $AA^T = U_r \Sigma_r^2 U_r^T$. [2]

iii It holds that

$$\|V_r^T q\|_2^2 = q^T V V^T q = q^T q = 1,$$

where the second equality comes from the fact that the columns of V are orthonormal. Furthermore,

$$\begin{aligned} \|V_r^T q\|_2^2 &= q^T V_r V_r^T q = \sum_{i=1}^r \langle q, v_i \rangle^2 \\ &\leq \sum_{i=1}^n \langle q, v_i \rangle^2 = q^T V V^T q = 1. \end{aligned}$$

[2]

iv Note that

$$\|Aq\|_2^2 = q^T A^T A q = q^T V \Sigma^2 V^T q = \sum_{i=1}^r \sigma_i^2 \langle v_i, q \rangle^2.$$

An upper bound can be obtained via

$$\|Aq\|_2^2 \leq \sum_{i=1}^r \sigma_{\max}^2 \langle v_i, q \rangle^2 = \sigma_{\max}^2 \left(\sum_{i=1}^r \langle v_i, q \rangle^2 \right) \leq \sigma_{\max}^2,$$

which can be achieved by picking q as v_1 the singular vector corresponding to σ_{\max} . Similarly, it holds that

$$\|Aq\|_2^2 \geq \sum_{i=1}^r \sigma_{\min}^2 \langle v_i, q \rangle^2 + \sum_{i=r+1}^n 0 \cdot \langle v_i, q \rangle^2 \geq 0,$$

which can be achieved when $q \in \text{span}\{v_{r+1}, \dots, v_n\}$. [4]

v $A^\dagger = V_r \Sigma_r^{-1} U_r^T$ and $AA^\dagger = U_r U_r^T$. [2]

vi The orthogonality can be verified as

$$\begin{aligned}
A^T \mathbf{x}_{\text{res}} &= A^T \mathbf{x} - A^T (A A^\dagger) \mathbf{x} = (A^T - A^T (A A^\dagger)) \mathbf{x} \\
&= (V_r \Sigma_r U_r^T - V_r \Sigma_r U_r^T (U_r U_r^T)) \mathbf{x} \\
&= (V_r \Sigma_r U_r^T - V_r \Sigma_r U_r^T) \mathbf{x} = 0.
\end{aligned}$$

[2]

(b)

i Let $\mathcal{I} \subset [n]$ and $|\mathcal{I}| \leq 2S$. For any submatrix of \mathbf{A} , say $\mathbf{A}_{\mathcal{I}}$, it holds that

$$\sqrt{1 - \delta_{2S}} \leq \sigma_{\min}(\mathbf{A}_{\mathcal{I}}) \leq \sigma_{\max}(\mathbf{A}_{\mathcal{I}}) \leq \sqrt{1 + \delta_{2S}}.$$

[2]

ii It holds that

$$\sqrt{1 - \delta_{2S}} \sqrt{2} \leq \|\mathbf{A}_{\mathcal{I}} \mathbf{p} + \mathbf{A}_{\mathcal{J}} \mathbf{q}\|_2 = \left\| [\mathbf{A}_{\mathcal{I}}, \mathbf{A}_{\mathcal{J}}] \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \right\|_2 \leq \sqrt{1 + \delta_{2S}} \sqrt{2}.$$

Similarly,

$$\sqrt{1 - \delta_{2S}} \sqrt{2} \leq \|\mathbf{A}_{\mathcal{I}} \mathbf{p} - \mathbf{A}_{\mathcal{J}} \mathbf{q}\|_2 = \left\| [\mathbf{A}_{\mathcal{I}}, \mathbf{A}_{\mathcal{J}}] \begin{bmatrix} \mathbf{p} \\ -\mathbf{q} \end{bmatrix} \right\|_2 \leq \sqrt{1 + \delta_{2S}} \sqrt{2}.$$

[2]

iii Note that

$$\begin{aligned}
4\mathbf{p}^T \mathbf{A}_{\mathcal{I}}^T \mathbf{A}_{\mathcal{J}} \mathbf{q} &= \|\mathbf{A}_{\mathcal{I}} \mathbf{p} + \mathbf{A}_{\mathcal{J}} \mathbf{q}\|_2^2 - \|\mathbf{A}_{\mathcal{I}} \mathbf{p} - \mathbf{A}_{\mathcal{J}} \mathbf{q}\|_2^2 \\
&\leq 2(1 + \delta_{2S}) - 2(1 - \delta_{2S}) = 4\delta_{2S}.
\end{aligned}$$

As \mathbf{p} and \mathbf{q} can be chosen arbitrarily, it follows $\sigma_{\max}(\mathbf{A}_{\mathcal{I}}^T \mathbf{A}_{\mathcal{J}}) \leq \delta_{2S}$. [2]

Solution to Problem 2.

(a)

$$d(\mathbf{x}, \mathcal{P}) = \frac{|\boldsymbol{\beta}^T \mathbf{x} - b|}{\|\boldsymbol{\beta}\|_2}.$$

[2]

(b) There exist $\boldsymbol{\beta}$ and b such that

$$\begin{aligned} \boldsymbol{\beta}^T \mathbf{x}_i + b &\geq +1 \quad \text{for } y_i = +1, \\ \boldsymbol{\beta}^T \mathbf{x}_i + b &\leq -1 \quad \text{for } y_i = -1. \end{aligned}$$

or equivalently

$$y_i (\boldsymbol{\beta}^T \mathbf{x}_i + b) - 1 \geq 0, \quad \forall i.$$

[2]

(c)

$$\begin{aligned} \min_{\boldsymbol{\beta}, b} \quad & \frac{1}{2} \|\boldsymbol{\beta}\|_2^2, \\ \text{subject to} \quad & 1 - y_i (\boldsymbol{\beta}^T \mathbf{x}_i + b) \leq 0. \end{aligned}$$

[2]

(d)

$$L(\boldsymbol{\beta}, b, \boldsymbol{\lambda}) = \frac{1}{2} \|\boldsymbol{\beta}\|_2^2 + \sum_i \lambda_i (1 - y_i (\boldsymbol{\beta}^T \mathbf{x}_i + b)),$$

where $\lambda_i \geq 0$, $i = 1, 2, \dots, m$.

[2]

(e) $L_D(\boldsymbol{\lambda}) = \min_{\boldsymbol{\beta}, b} L(\boldsymbol{\beta}, b, \boldsymbol{\lambda})$. To find its specific form, we compute $\partial L(\boldsymbol{\beta}, b, \boldsymbol{\lambda}) / \partial \boldsymbol{\beta}$ and $\partial L(\boldsymbol{\beta}, b, \boldsymbol{\lambda}) / \partial b$, and set them to zero. In particular,

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} - \sum_i \lambda_i y_i \mathbf{x}_i = 0 \Rightarrow \boldsymbol{\beta} = \sum_i \lambda_i y_i \mathbf{x}_i,$$

and

$$\frac{\partial L}{\partial b} = \sum_i \lambda_i y_i = 0.$$

Hence,

$$\begin{aligned}
L_D &= \frac{1}{2} \|\beta\|_2^2 + \sum_i \lambda_i - \beta^T \left(\sum_i \lambda_i y_i \mathbf{x}_i \right) + b \sum_i \lambda_i y_i \\
&= \sum_i \lambda_i - \frac{1}{2} \beta^T \beta + 0 \\
&= -\frac{1}{2} \lambda^T K \lambda + \mathbf{1}^T \lambda,
\end{aligned}$$

where $K_{i,j} = y_i \mathbf{x}_i^T \mathbf{x}_j y_j$.

[4]

(f) The dual problem is given by

$$\begin{aligned}
&\max_{\lambda} \quad -\frac{1}{2} \lambda^T K \lambda + \mathbf{1}^T \lambda, \\
&\text{subject to} \quad \lambda_i \geq 0, \quad \forall i, \\
&\quad \quad \quad \sum_i \lambda_i y_i = 0.
\end{aligned}$$

[2]

(g) The KKT conditions are given by

$$\begin{aligned}
\frac{\partial L}{\partial \beta} &= \beta - \sum_i \lambda_i y_i \mathbf{x}_i = 0, \\
\frac{\partial L}{\partial b} &= \sum_i \lambda_i y_i = 0, \\
1 - y_i (\beta^T \mathbf{x}_i + b) &\leq 0, \\
\lambda_i &\geq 0, \\
\lambda_i (1 - y_i (\beta^T \mathbf{x}_i + b)) &= 0.
\end{aligned}$$

[4]

(h) The last condition in the KKT conditions implies that

$$\begin{cases} \text{if } \lambda_i \neq 0 & \text{then } 1 = y_i (\beta^T \mathbf{x}_i + b), \\ \text{if } 1 \neq y_i (\beta^T \mathbf{x}_i + b) & \text{then } \lambda_i = 0. \end{cases}$$

Note that the optimal β is given by

$$\beta = \sum_i \lambda_i y_i \mathbf{x}_i,$$

where only the λ_i 's that $\lambda_i > 0$ matter. Hence,

$$\boldsymbol{\beta} = \sum_{i \in \mathcal{I}} \lambda_i y_i \mathbf{x}_i, \quad \mathcal{I} = \{i : y_i (\boldsymbol{\beta}^T \mathbf{x}_i + b) = 1\}.$$

[2]

Solution to Problem 3.

(a)

i

$$\frac{d|x|}{dx} = \begin{cases} 1 & \text{if } x > 0, \\ [-1, 1] & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases} \quad [2]$$

ii Since $f(\mathbf{x}) = \sum_i \frac{1}{2} (x_i - z_i)^2 + \lambda |\mathbf{x}|$, the optimal \mathbf{x}^* is given by

$$x_i^* = \eta(z_i; \lambda) = \begin{cases} z_i - \lambda & \text{if } z_i \geq \lambda, \\ 0 & \text{if } -\lambda < z_i < \lambda, \\ z_i + \lambda & \text{if } z_i \leq -\lambda. \end{cases} \quad [2]$$

iii It is clear that $\nabla g(\mathbf{x}) = -\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x})$. [2]

iv

$$\begin{aligned} \tilde{g}(\mathbf{x}) &= g(\mathbf{x}^k) + \langle \mathbf{x} - \mathbf{x}^k, \nabla g(\mathbf{x}^k) \rangle + \frac{1}{2t_k} \|\mathbf{x} - \mathbf{x}^k\|_2^2 \\ &= \frac{1}{2t_k} m \left(\|\mathbf{x} - \mathbf{x}^k\|_2^2 + 2 \langle \mathbf{x} - \mathbf{x}^k, t_k \nabla g(\mathbf{x}^k) \rangle + 2t_k g(\mathbf{x}^k) \right) \\ &= \frac{1}{2t_k} \left(\|\mathbf{x} - \mathbf{x}^k + t_k \nabla g(\mathbf{x}^k)\|_2^2 + 2t_k c_2 \right) \\ &= \frac{1}{2t_k} \|\mathbf{x} - (\mathbf{x}^k - t_k \nabla g(\mathbf{x}^k))\|_2^2 + c_2. \end{aligned}$$

As a result, $c_1 = 2t_k$ and $\mathbf{z} = \mathbf{x}^k - t_k \nabla g(\mathbf{x}^k) = \mathbf{x}^k + t_k \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^k)$. [3]

v The minimum of

$$\begin{aligned} \tilde{g}(\mathbf{x}) + \lambda \|\mathbf{x}\|_1 &= \frac{1}{2t_k} \|\mathbf{x} - (\mathbf{x}^k + t_k \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^k))\|_2^2 + \lambda \|\mathbf{x}\|_1 + c_2 \\ &= \frac{1}{t_k} \left[\frac{1}{2} \|\mathbf{x} - (\mathbf{x}^k + t_k \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^k))\|_2^2 + \lambda t_k \|\mathbf{x}\|_1 \right] + c_2 \end{aligned}$$

is achieved by \mathbf{x}^* with

$$x_i^* = \eta((\mathbf{x}^k + t_k \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^k))_i; \lambda t_k).$$

Hence the ISTA is given by

$$\mathbf{x}^k = \eta (\mathbf{x}^{k-1} + t_k \mathbf{A}^T (\mathbf{y} - \mathbf{A} \mathbf{x}^{k-1}) ; \lambda t_k).$$

[3]

(b)

i

$$\begin{aligned} \|\mathbf{A}\mathbf{h}\|_2 &= \|\mathbf{A}(\mathbf{x}^\# - \mathbf{x}_0)\|_2 = \|\mathbf{A}\mathbf{x}^\# - \mathbf{y} + \mathbf{y} - \mathbf{A}\mathbf{x}_0\|_2 \\ &\leq \|\mathbf{y} - \mathbf{A}\mathbf{x}^\#\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}_0\|_2 \\ &\leq \epsilon + \epsilon = 2\epsilon, \end{aligned}$$

where the first inequality follows from triangular inequality, and the second one holds due to the assumptions. [2]

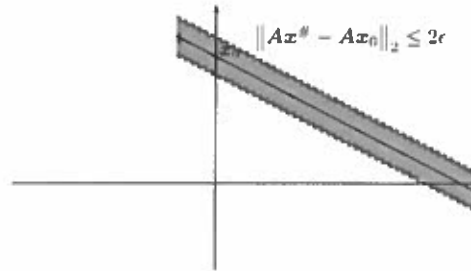
ii As $\mathbf{x}^\#$ is the output of ℓ_1 -minimization,

$$\begin{aligned} \|\mathbf{x}_0\|_1 &\geq \|\mathbf{x}^\#\|_1 = \|\mathbf{x}_0 + \mathbf{h}\|_1 \\ &= \|(\mathbf{x}_0 + \mathbf{h})_{\mathcal{T}_0}\|_1 + \|\mathbf{h}_{\mathcal{T}_0^c}\|_1 \\ &\geq \|\mathbf{x}_0\|_1 - \|\mathbf{h}_{\mathcal{T}_0}\|_1 + \|\mathbf{h}_{\mathcal{T}_0^c}\|_1. \end{aligned}$$

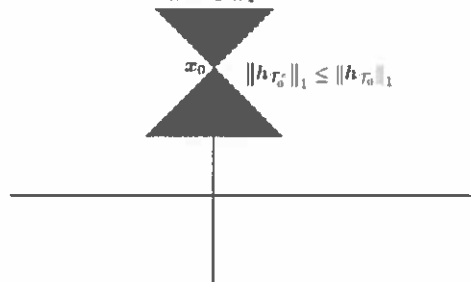
Hence $\|\mathbf{h}_{\mathcal{T}_0^c}\|_1 \leq \|\mathbf{h}_{\mathcal{T}_0}\|_1$.

[3]

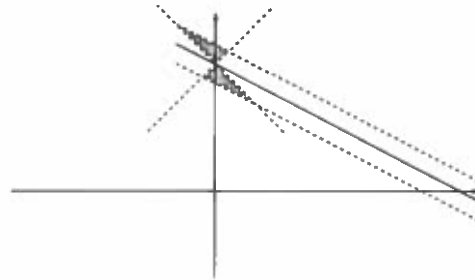
iii The region defined by $\|\mathbf{A}\mathbf{h}\|_2 \leq 2\epsilon$ is given by



The region corresponding to $\|\mathbf{h}_{\mathcal{T}_0^c}\|_1 \leq \|\mathbf{h}_{\mathcal{T}_0}\|_1$ is given by



The intersection of these two regions determines a small h :



[3]

Solution to Problem 3.

(a) The complete \mathbf{X} is given by

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

From the first four elements in the 1st and 3rd rows, these two rows are linearly independent. The difference between these two subvectors is $[1, 1, 1, 1]$, which leads to the guess that likely the first four elements in the last row should be 1, 1, 1, 1. Once having accomplished this, one should be able to figure out the rest straightforwardly using the information $\text{rank}(\mathbf{X}) = 2$. [5]

(b)

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{H}\Sigma_x\mathbf{H}^T + \Sigma_w).$$

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Sigma_x & \Sigma_x\mathbf{H}^T \\ \mathbf{H}\Sigma_x & \mathbf{H}\Sigma_x\mathbf{H}^T + \Sigma_w \end{bmatrix}\right).$$

To estimate \mathbf{X} from the given observation $\mathbf{Y} = \mathbf{y}$, we can use Gaussian conditioning lemma. Let

$$\Sigma = \begin{bmatrix} \Sigma_x & \Sigma_x\mathbf{H}^T \\ \mathbf{H}\Sigma_x & \mathbf{H}\Sigma_x\mathbf{H}^T + \Sigma_w \end{bmatrix}.$$

Find its inverse $\mathbf{K} := \Sigma^{-1}$. Apply Gaussian conditioning lemma to compute the conditional mean $\mathbb{E}[\mathbf{X}|\mathbf{Y} = \mathbf{y}]$ and use it as the estimate of \mathbf{X} . [5]

ii If nodes a and b do not have a connection, then X_a and X_b are independent conditioned on $\mathbf{X}_{\sim\{a,b\}}$. In other words, a and b have a connection when X_a and X_b are correlated conditioned on $\mathbf{X}_{\sim\{a,b\}}$. Based on Gaussian conditioning lemma, the covariance matrix of $\mathbf{X}_{\{a,b\}}|\mathbf{X}_{\sim\{a,b\}}$ is given by $\mathbf{K}_{\{a,b\},\{a,b\}}^{-1}$. One simply needs to check whether the off-diagonal element of the matrix $\mathbf{K}_{\{a,b\},\{a,b\}}$ is zero or not: if it is zero, nodes a and b have no connection; otherwise nodes a and b have a connection. [5]

iii The expectation can be approximated by the average:

$$\begin{aligned}
& \mathbb{E} \left[\sum_a \left(X_a - \sum_{b \neq a} \theta_{ba} X_b \right)^2 \right] \\
& \approx \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_{(i)} - \Theta^T \mathbf{x}_{(i)})^T (\mathbf{x}_{(i)} - \Theta^T \mathbf{x}_{(i)}) \\
& = \frac{1}{m} \left\| \begin{bmatrix} \mathbf{x}_{(1)}^T \\ \vdots \\ \mathbf{x}_{(m)}^T \end{bmatrix} - \begin{bmatrix} \mathbf{x}_{(1)}^T \\ \vdots \\ \mathbf{x}_{(m)}^T \end{bmatrix} \Theta \right\|_F^2 \\
& = \frac{1}{m} \|\mathbf{X} - \mathbf{X}\Theta\|_F^2.
\end{aligned}$$

Based on it, the optimization problem is given by

$$\min_{\Theta \in \Theta} \frac{1}{m} \|\mathbf{X} - \mathbf{X}\Theta\|_F^2 + \lambda \sum_{a \neq b} |\theta_{ab}|,$$

where $\Theta = \{\Theta : \text{diag}(\Theta) = \mathbf{0}\}$.

[5]