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

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

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

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

[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 $\sigma_{\max}$. 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 $q \in \operatorname{span}\{v_{r+1}, \cdots, v_n\}$. [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

$$\begin{aligned} \boldsymbol{A}^T\boldsymbol{x}_{\mathrm{res}} &= \boldsymbol{A}^T\boldsymbol{x} - \boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{A}^\dagger\right)\boldsymbol{x} = \left(\boldsymbol{A}^T - \boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{A}^\dagger\right)\right)\boldsymbol{x} \\ &= \left(\boldsymbol{V}_r\boldsymbol{\Sigma}_r\boldsymbol{U}_r^T - \boldsymbol{V}_r\boldsymbol{\Sigma}_r\boldsymbol{U}_r^T\left(\boldsymbol{U}_r\boldsymbol{U}_r^T\right)\right)\boldsymbol{x} \\ &= \left(\boldsymbol{V}_r\boldsymbol{\Sigma}_r\boldsymbol{U}_r^T - \boldsymbol{V}_r\boldsymbol{\Sigma}_r\boldsymbol{U}_r^T\right)\boldsymbol{x} = \boldsymbol{0}. \end{aligned}$$

[2]

(b)

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

$$\sqrt{1-\delta_{2S}} \le \sigma_{\min}\left(\boldsymbol{A}_{\mathcal{I}}\right) \le \sigma_{\max}\left(\boldsymbol{A}_{\mathcal{I}}\right) \le \sqrt{1+\delta_{2S}}.$$

[2]

ii It holds that

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

Similarly,

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

[2]

iii Note that

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

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

Solution to Problem 2.

(a)

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

[2]

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

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

or equivalently

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

[2]

(c)

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

[2]

(d)

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

where $\lambda_i \geq 0$, $i = 1, 2, \cdots, m$. [2]

(e) $L_D(\boldsymbol{\lambda}) = \min\limits_{\boldsymbol{\beta}, b} L(\boldsymbol{\beta}, b, \boldsymbol{\lambda})$. 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,

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

and

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

Hence,

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

where $K_{i,j} = y_i \boldsymbol{x}_i^T \boldsymbol{x}_j y_j$. [4]

(f) The dual problem is given by

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

[2]

(g) The KKT conditions are given by

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

[4]

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

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

Note that the optimal $\boldsymbol{\beta}$ is given by

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

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

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

[2]

Solution to Problem 3.

(a)

i

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

[2]

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

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

[2]

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

iv

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

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

v The minimum of

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

is achieved by $\boldsymbol{x}^*$ with

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

Hence the ISTA is given by

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

[3]

(b)

i

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

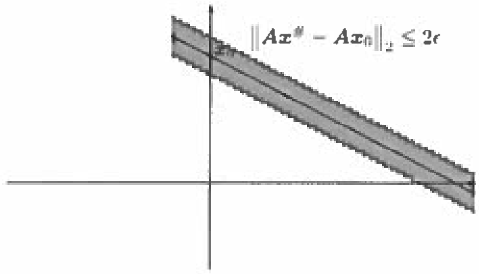
where the first inequality follows from triangular inequality, and the second one holds due to the assumptions. [2]

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

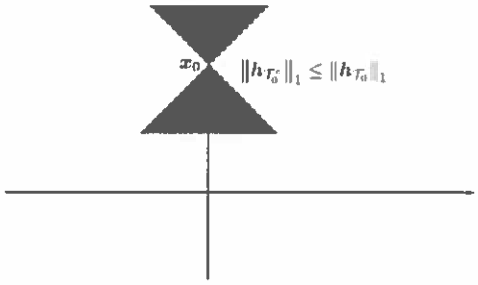
$$\begin{aligned}\|\boldsymbol{x}_0\|_1 \geq \left\|\boldsymbol{x}^\#\right\|_1 &= \|\boldsymbol{x}_0 + \boldsymbol{h}\|_1 \\ &= \left\|(\boldsymbol{x}_0 + \boldsymbol{h})_{\mathcal{T}_0}\right\|_1 + \left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1 \\ &\geq \|\boldsymbol{x}_0\|_1 - \|\boldsymbol{h}_{\mathcal{T}_0}\|_1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

Hence $\left\|\boldsymbol{h}_{\mathcal{T}_0^c}\right\|_1 \leq \|\boldsymbol{h}_{\mathcal{T}_0}\|_1$. [3]

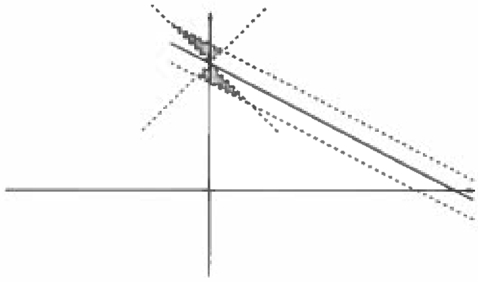
iii The region defined by $\|\boldsymbol{A}\boldsymbol{h}\|_2 \leq 2\epsilon$ is given by



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



The intersection of these two regions determines a small $\boldsymbol{h}$:



[3]

Solution to Problem 3.

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

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

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 $\operatorname{rank}(\boldsymbol{X}) = 2$. [5]

(b)

i $\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_w\right)$.

$$\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x\boldsymbol{H}^T \\ \boldsymbol{H}\boldsymbol{\Sigma}_x & \boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_w \end{bmatrix}\right).$$

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

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x\boldsymbol{H}^T \\ \boldsymbol{H}\boldsymbol{\Sigma}_x & \boldsymbol{H}\boldsymbol{\Sigma}_x\boldsymbol{H}^T + \boldsymbol{\Sigma}_w \end{bmatrix}.$$

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

ii If nodes a and b do not have a connection, then X_a and X_b are independent conditioned on $\boldsymbol{X}_{\sim\{a,b\}}$. In other words, a and b have a connection when X_a and X_b are correlated conditioned on $\boldsymbol{X}_{\sim\{a,b\}}$. Based on Gaussian conditioning lemma, the covariance matrix of $\boldsymbol{X}_{\{a,b\}} | \boldsymbol{X}_{\sim\{a,b\}}$ is given by $\boldsymbol{K}^{-1}_{\{a,b\},\{a,b\}}$. One simply needs to check whether the off-diagonal element of the matrix $\boldsymbol{K}_{\{a,b\},\{a,b\}}$ 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:

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

Based on it, the optimization problem is given by

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

where $\Theta = \{\Theta : \ \mathrm{diag}(\Theta) = \mathbf{0}\}$. [5]