Solution to Problem 1.

(a)

i A vector $\boldsymbol{v} \in \mathbb{R}^n$ is an eigenvector of $\boldsymbol{M}$ if $\boldsymbol{M}\boldsymbol{v} = \lambda \boldsymbol{v}$ where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue. [1]

ii The i^{th} eigenvector and eigenvalue of $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{A}^T$ are given by $\boldsymbol{u}_i$ and σ_i^2 respectively. This can be justified by

$$\begin{aligned} \boldsymbol{B}\boldsymbol{u}_i &= \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{u}_i = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{u}_i \\ &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{u}_i = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{e}_i \\ &= \sigma_i^2\boldsymbol{U}\boldsymbol{e}_i = \sigma_i^2\boldsymbol{u}_i, \end{aligned}$$

where $\boldsymbol{e}_i$ is the standard basis vector. [2]

iii Let r be the rank of $\boldsymbol{A}$. Define $\boldsymbol{U}_r = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r]$, $\boldsymbol{\Sigma}_r = \operatorname{diag}([\sigma_1, \cdots, \sigma_r])$, and $\boldsymbol{V}_r = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r]$. Then $\boldsymbol{A}^\dagger = \boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}\boldsymbol{U}_r^T$. [1]

iv $\operatorname{proj}(\boldsymbol{x}, \boldsymbol{A}) = \boldsymbol{A}\boldsymbol{A}^\dagger\boldsymbol{x}$. [1]

v The orthogonality can be verified as

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

[2]

vi Note that

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{v}\|_2^2 &= \|\boldsymbol{x} - \boldsymbol{x}_p + \boldsymbol{x}_p - \boldsymbol{v}\|_2^2 = \|\boldsymbol{x}_r + \boldsymbol{x}_p - \boldsymbol{v}\|_2^2 \\ &= \|\boldsymbol{x}_r\|_2^2 + 2\boldsymbol{x}_r^T\left(\boldsymbol{x}_p - \boldsymbol{v}\right) + \|\boldsymbol{x}_p - \boldsymbol{v}\|_2^2 \\ &= \|\boldsymbol{x}_r\|_2^2 + \|\boldsymbol{x}_p - \boldsymbol{v}\|_2^2, \end{aligned}$$

where the last equality comes from the orthogonality between $\boldsymbol{x}_r$ and $\boldsymbol{x}_p - \boldsymbol{v}$. Since $\|\boldsymbol{x}_p - \boldsymbol{v}\|_2^2 \geq 0$, it is clear that $\|\boldsymbol{x} - \boldsymbol{v}\|_2 \geq \|\boldsymbol{x}_r\|_2$. [3]

(b)

i

$$\begin{aligned}\operatorname{tr}(\boldsymbol{AB}) &= \sum_i (\boldsymbol{AB})_{i,i} = \sum_i \sum_j A_{i,j} B_{j,i} \\ &= \sum_j \sum_i B_{j,i} A_{i,j} = \sum_j (\boldsymbol{BA})_{j,j} = \operatorname{tr}(\boldsymbol{BA}).\end{aligned}$$

[2]

ii

$$\begin{aligned}\|\boldsymbol{A}\|_F^2 &= \sum_{i,j} A_{i,j}^2 = \operatorname{tr}\left(\boldsymbol{A}^T\boldsymbol{A}\right) = \operatorname{tr}\left(\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T\right) \\ &= \operatorname{tr}\left(\boldsymbol{\Sigma}^2\boldsymbol{V}^T\boldsymbol{V}\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}^2\right) = \sum_i \sigma_i^2.\end{aligned}$$

[2]

iii $\|\boldsymbol{A}\|_2 = \sigma_{\max} = \sigma_1$. This can be proved as follows.

$$\begin{aligned}\|\boldsymbol{Ax}\|_2^2 &= \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{x}^T\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T\boldsymbol{x} \\ &= \sum_i \sigma_i^2 \left(\boldsymbol{x}^T\boldsymbol{v}_i\right)^2 \\ &\le \sigma_1^2 \sum_i \left(\boldsymbol{x}^T\boldsymbol{v}_i\right)^2 = \sigma_1^2\boldsymbol{x}^T\boldsymbol{V}\boldsymbol{V}^T\boldsymbol{x} \\ &= \sigma_1^2\boldsymbol{x}^T\boldsymbol{x} = \sigma_1^2.\end{aligned}$$

[3]

iv Since $\boldsymbol{A} = \boldsymbol{A}^T$ and $\boldsymbol{A} \ge 0$, the singular value decomposition of $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T$. Then

$$\begin{aligned}\operatorname{tr}(\boldsymbol{A}) &= \operatorname{tr}\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}\boldsymbol{U}^T\boldsymbol{U}\right) = \operatorname{tr}(\boldsymbol{\Sigma}) \\ &= \sum_{i=1}^{\min(m,n)} \sigma_i = \|\boldsymbol{A}\|_*.\end{aligned}$$

[3]

Solution to Problem 2.

(a)

i

A. The soft-thresholding function is of the form

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

[1]

B. The IST algorithm is an iterative algorithm where in the k^{th} iteration the variable $\boldsymbol{x}^k$ is updated by

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

and $t_k > 0$ is an appropriately chosen step size. [1]

ii

A. Given an input vector $\boldsymbol{z}$, the hard thresholding function $H_S(\boldsymbol{z})$ sets all but the largest (in magnitude) S elements of $\boldsymbol{z}$ to zero. It is designed to solve the non-convex optimisation problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 \text{ subject to } \|\boldsymbol{x}\|_0 \leq S.$$

[2]

B. The IHT algorithm is an iterative algorithm where in the k^{th} iteration the variable $\boldsymbol{x}^k$ is updated by

$$\boldsymbol{x}^k = H_S\left(\boldsymbol{x}^{k-1} + t_k \boldsymbol{A}^T\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{k-1}\right)\right),$$

and $t_k > 0$ is an appropriately chosen step size. It is designed to solve the non-convex optimisation problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \text{ subject to } \|\boldsymbol{x}\|_0 \leq S.$$

[2]

(b)

i Define the soft thresholding function $\eta_\sigma(\boldsymbol{Z};\lambda)$ as

$$\boldsymbol{X} = \eta_\sigma(\boldsymbol{Z};\lambda) = \sum_i \boldsymbol{u}_i \eta(\sigma_i;\lambda)\boldsymbol{v}_i^T,$$

where σ_i is the i^{th} singular value of $\boldsymbol{Z}$, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the corresponding singular vectors. The IST algorithm to solve the low-rank matrix recovery problem is an iterative algorithm where in the k^{th} iteration the matrix $\boldsymbol{X}^k$ is updated by

$$\boldsymbol{X}^k = \eta_\sigma\left(\boldsymbol{X}^{k-1} + t_k\mathcal{A}^*\left(\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{X}^{k-1}\right)\right); \lambda t_k\right),$$

and $t_k > 0$ is an appropriately chosen step size. [2]

ii Define the hard thresholding function $H_r(\boldsymbol{Z};\lambda)$ as

$$\boldsymbol{X} = H_r(\boldsymbol{Z}) = \boldsymbol{U}_r \mathrm{diag}\left([\sigma_1,\cdots,\sigma_r]\right)\boldsymbol{V}_r^T,$$

where σ_i is the i^{th} singular value of $\boldsymbol{Z}$, $\boldsymbol{U}_r = [\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r]$ and $\boldsymbol{V}_r = [\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r]$. The IHT algorithm to solve the low-rank matrix recovery problem is an iterative algorithm where in the k^{th} iteration the matrix $\boldsymbol{X}^k$ is updated by

$$\boldsymbol{X}^k = H_r\left(\boldsymbol{X}^{k-1} + t_k\mathcal{A}^*\left(\boldsymbol{y} - \mathcal{A}\left(\boldsymbol{X}^{k-1}\right)\right)\right),$$

and $t_k > 0$ is an appropriately chosen step size. [2]

iii

A. Since $\boldsymbol{X}$ is of rank 1, $\boldsymbol{X} = \boldsymbol{u}\boldsymbol{v}^T$ and $\boldsymbol{u}$ must be of the form $\boldsymbol{u} = [1,1,-2]^T$. Hence $\boldsymbol{v} = [1,-1,-2]^T$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 & -2 \\ 1 & -1 & -2 \\ -2 & 2 & 4 \end{bmatrix}.$$

[3]

B.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$\boldsymbol{y} = [1, 1, -1, 2, 4]^T$, and

$$\mathcal{A}^* (\boldsymbol{y}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 2 & 4 \end{bmatrix}.$$

[3]

(c) Write

$$\boldsymbol{y} = \mathcal{A}\left(\begin{bmatrix} h(1) \\ h(2) \\ h(3) \end{bmatrix} \begin{bmatrix} x(1) \\ x(2) \\ x(3) \end{bmatrix}^T \right) = \mathcal{A}\left(\begin{bmatrix} h(1)x(1) & h(1)x(2) & h(1)x(3) \\ h(2)x(1) & h(2)x(2) & h(2)x(3) \\ h(3)x(1) & h(3)x(2) & h(3)x(3) \end{bmatrix} \right).$$

Clearly the matrix presentation of $\mathcal{A}$ is given by

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Possible convex optimisation formulations for blind deconvolution include

$$\min_{\boldsymbol{X}} \frac{1}{2} \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_2^2 + \lambda \|\boldsymbol{X}\|_* ,$$

and

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_* \text{ subject to } \boldsymbol{y} = \mathcal{A}(\boldsymbol{X}).$$

[4]

Solution to Problem 3.

(a)

i A set $S \subset \mathbb{R}^n$ is convex if

$$\lambda \boldsymbol{x} + (1-\lambda)\,\boldsymbol{y} \in S,$$

for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $\lambda \in [0,1]$. [2]

ii A function $f:\ \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(\lambda \boldsymbol{x} + (1-\lambda)\,\boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + (1-\lambda)\, f(\boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and all $\lambda \in [0,1]$. [2]

iii The standard form of a convex optimisation problem is

$$\begin{aligned} &\min_{\boldsymbol{x}}\ f(\boldsymbol{x}) \\ &\text{subject to } h_i(\boldsymbol{x}) \le 0, \quad i \in \{1, \cdots, m\} \\ &\qquad\qquad \ell_j(\boldsymbol{x}) = 0, \quad j \in \{1, \cdots, p\} \end{aligned}$$

where $f(\boldsymbol{x})$, $h_i(\boldsymbol{x})$ are convex and ℓ_j is affine. [2]

iv The corresponding Lagrangian is given by

$$L(\boldsymbol{x}, \boldsymbol{u}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{u}^T \boldsymbol{A} \boldsymbol{x}.$$

Minimise the Lagrangian with respect to $\boldsymbol{x}$ gives

$$\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{u} = \boldsymbol{0}.$$

Combine this with the constraints $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. One has

$$\begin{bmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{c} \\ \boldsymbol{0} \end{bmatrix}.$$

Hence the optimal $\boldsymbol{x}$ can be computed by

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}^{\dagger} \begin{bmatrix} -\boldsymbol{c} \\ \boldsymbol{0} \end{bmatrix}$$

[3]

(b)

i For any $\boldsymbol{x}, \boldsymbol{y} \in C_\alpha$, it holds that $\forall \lambda \in [0, 1]$,

$$\begin{aligned} f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &\leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \\ &\leq \lambda \alpha + (1-\lambda)\alpha = \alpha. \end{aligned}$$

Hence $\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in C_\alpha$. [2]

ii The sublevel set $C_0 = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2 = \sqrt{y} \right\}$ which gives the shell of the ball with radius $\sqrt{y}$. Pick arbitrary two points on the shell. The line segment between these two points is not on the shell (in the ball defined by the shell). Hence the sublevel set C_0 is not convex and f is not convex. [2]

(c)

i Note that $\left(\boldsymbol{a}_i^T \boldsymbol{x}\right)^2 = \left(\boldsymbol{a}_i^T \boldsymbol{x}\right)\left(\boldsymbol{x}^T \boldsymbol{a}_i\right) = \boldsymbol{a}_i^T \boldsymbol{x}\left(\boldsymbol{x}^T \boldsymbol{a}_i\right) = \operatorname{tr}\left(\boldsymbol{x}\left(\boldsymbol{x}^T \boldsymbol{a}_i\right)\boldsymbol{a}_i^T\right) = \operatorname{tr}(\boldsymbol{X}\boldsymbol{A}_i)$. One has $f(\boldsymbol{x}) = \sum_i \left(y_i - \left(\boldsymbol{a}_i^T \boldsymbol{x}\right)^2\right)^2 = \sum_i \left(y_i - \operatorname{tr}(\boldsymbol{X}\boldsymbol{A}_i)\right)^2$. [2]

ii

A. It is sufficient to show that the objective function is convex and the constraint set is convex.

The first term in the objective function is a quadratic function of $\boldsymbol{X}$ and hence convex. In fact, it can be written as $\sum_i \left(y_i - \mathcal{A}(\boldsymbol{X})\right)^2$ where $\mathcal{A}$ is a linear operator. The second term in the objective function is a linear function of $\boldsymbol{X}$ and hence convex.

The first constraint is essentially to say $x_{i,j} - x_{j,i} = 0$. It is an equality constraint involving linear functions of $\boldsymbol{X}$ and hence defines a convex set.

The only non-trivial part is to verify that the constraint $\boldsymbol{X} \geq 0$ gives a convex set. To show that, note that $\boldsymbol{X} \geq 0$ if and only if $\boldsymbol{v}^T \boldsymbol{X} \boldsymbol{v} \geq 0$ for all $\boldsymbol{v}$. Let $\boldsymbol{X}_1 \geq 0$ and $\boldsymbol{X}_2 \geq 0$. Then for all $\lambda \in [0, 1]$,

$$\begin{aligned} \boldsymbol{v}^T\left(\lambda \boldsymbol{X}_1 + (1-\lambda)\boldsymbol{X}_2\right)\boldsymbol{v} &= \lambda \boldsymbol{v}^T \boldsymbol{X}_1 \boldsymbol{v} + (1-\lambda)\boldsymbol{v}^T \boldsymbol{X}_2 \boldsymbol{v} \\ &\geq 0. \end{aligned}$$

Hence the constraint $\boldsymbol{X} \geq 0$ gives a convex set. [3]

B. The goal of using the optimisation problem (3.4) is to find an $\boldsymbol{X}$

such that $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^T$. The first term in the objective function is for data consistency. The second term in the objective function is to promote low-rank matrices as solutions: the matrix $\boldsymbol{x}\boldsymbol{x}^T$ has rank 1 and $\mathrm{tr}(\boldsymbol{X}) = \|\boldsymbol{X}\|_*$ under the constraints given in (3.4). The two constraints are motivated by the fact that the matrix $\boldsymbol{x}\boldsymbol{x}^T$ is symmetric and non-negative definite. [2]

Solution to Problem 4.

(a) The mutual coherence is defined as

$$\mu(\boldsymbol{A}) = \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle| = \max_{i \neq j} \left| \boldsymbol{a}_i^T \boldsymbol{a}_j \right|.$$

[1]

(b)

i A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is said to satisfy the RIP with parameters (K, δ), if for all $\mathcal{T} \subset \{1, \cdots, n\}$ such that $|\mathcal{T}| \leq K$ and for all $\boldsymbol{q} \in \mathbb{R}^{|\mathcal{T}|}$, it holds that

$$(1-\delta)\|\boldsymbol{q}\|_2^2 \leq \|\boldsymbol{A}_{\mathcal{T}}\boldsymbol{q}\|_2^2 \leq (1+\delta)\|\boldsymbol{q}\|_2^2.$$

The RIC δ_K is defined as the smallest constant δ for which the (K, δ)-RIP holds, i.e.,

$$\delta_K = \inf \left\{ \delta : (1-\delta)\|\boldsymbol{q}\|_2^2 \leq \|\boldsymbol{A}_{\mathcal{T}}\boldsymbol{q}\|_2^2 \leq (1+\delta)\|\boldsymbol{q}\|_2^2 \right.$$
$$\left. \forall |\mathcal{T}| \leq K, \ \forall \boldsymbol{q} \in \mathbb{R}^{|\mathcal{T}|} \right\}.$$

[3]

ii It holds that

$$|\langle \boldsymbol{A}_{\mathcal{I}}\boldsymbol{a}, \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b} \rangle| \leq \delta_{k+\ell}\|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2,$$

that is, $c = \delta_{k+\ell}$. [1]

iii Suppose that there exists another S-sparse vector $\boldsymbol{x}' \neq \boldsymbol{x}_0$ such that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}'$. Then

$$\boldsymbol{A}(\boldsymbol{x}_0 - \boldsymbol{x}') = \boldsymbol{y} - \boldsymbol{y} = \boldsymbol{0}.$$

Note that $\boldsymbol{x}_0 - \boldsymbol{x}'$ has sparsity level at most $2S$. By the definition of RIP,

$$\|\boldsymbol{A}(\boldsymbol{x}_0 - \boldsymbol{x}')\|_2 \geq (1 - \delta_{2S})\|\boldsymbol{x}_0 - \boldsymbol{x}'\|_2 > 0.$$

We have a contradiction. The S-sparse solution of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ must be unique. [3]

(c) The diagonal elements of $\boldsymbol{M}$ are 1 and off-diagonal elements are bounded by μ by the definition of mutual coherence constant, i.e., $|M_{i,j}| < \mu$ for $i \neq j$.

By the Bershgorin circle theorem,

$$\lambda(M) \in \left[1 - \sum_{j \neq i} |M_{i,j}|, 1 + \sum_{j \neq i} |M_{i,j}|\right] \subset [1 - S\mu, 1 + S\mu].$$

Hence,

$$\delta_S \leq S\mu.$$

[3]

(d)

i Suppose that $i \notin \mathcal{T}$.

$$\begin{aligned} \left|\boldsymbol{a}_i^T \boldsymbol{y}\right| &= \left|\boldsymbol{a}_i^T \boldsymbol{A}\boldsymbol{x}\right| = \left|\boldsymbol{a}_i^T \boldsymbol{A}_{\mathcal{T}} \boldsymbol{x}_{0,\mathcal{T}}\right| \\ &\leq \delta_{S+1} \left\|\boldsymbol{x}_{0,\mathcal{T}}\right\|_2 = \delta_{S+1} \left\|\boldsymbol{x}_0\right\|_2. \end{aligned}$$

[3]

ii Suppose that $i \in \mathcal{T}$.

$$\begin{aligned} \left|\boldsymbol{a}_i^T \boldsymbol{y}\right| &= \left|\boldsymbol{a}_i^T \boldsymbol{a}_i x_{0,i} + \boldsymbol{a}_i^T \sum_j \boldsymbol{a}_j x_{0,j}\right| \\ &\geq \left|x_{0,i}\right| - \left|\boldsymbol{a}_i^T \boldsymbol{A}_{\mathcal{T}\setminus\{i\}} \boldsymbol{x}_{0,\mathcal{T}\setminus\{i\}}\right| \\ &\geq \left|x_{0,i}\right| - \delta_{S+1} \left\|\boldsymbol{x}_{0,\mathcal{T}\setminus\{i\}}\right\|_2 \\ &\geq \left|x_{0,i}\right| - \delta_{S+1} \left\|\boldsymbol{x}_0\right\|_2. \end{aligned}$$

At the same time,

$$\max_{i \in \mathcal{T}} \left|x_{0,i}\right| \geq \frac{1}{\sqrt{S}} \left\|\boldsymbol{x}_0\right\|.$$

Hence

$$\max_{i \in \mathcal{T}} \left|\boldsymbol{a}_i^T \boldsymbol{y}\right| \geq \left(\frac{1}{\sqrt{S}} - \delta_{S+1}\right) \left\|\boldsymbol{x}_0\right\|_2.$$

[3]

iii To guarantee that $i^* \in \mathcal{T}$, one needs

$$\frac{1}{\sqrt{S}} - \delta_{S+1} > \delta_{S+1}.$$

Or equivalently

$$\delta_{S+1} \leq \frac{1}{2\sqrt{S}}.$$

[3]