Dictionary Learning and Compressed Sensing

Gregor Bachmann, Leonard Adolphs, Emilien Pilloud

Administrative

- **No** exercise class next week!

Administrative

- **No** exercise class next week!
- Q & A session **next** Friday, May 29, 8:15-10am.

Notations

- $$\mathbf{A} = \begin{bmatrix} | & | & & | \\ \mathbf{A}_{\bullet 1} & \mathbf{A}_{\bullet 2} & \dots & \mathbf{A}_{\bullet N} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} — & \mathbf{A}_{1\bullet}^T & — \\ — & \mathbf{A}_{2\bullet}^T & — \\ & \vdots & \\ — & \mathbf{A}_{D\bullet}^T & — \end{bmatrix}$$

  where $\mathbf{A} \in \mathbb{R}^{D \times N}$, $\mathbf{A}_{i\bullet} \in \mathbb{R}^N$ and $\mathbf{A}_{\bullet i} \in \mathbb{R}^D$

  Warning: We will view both columns $\mathbf{A}_{\bullet l}$ and rows $\mathbf{A}_{l\bullet}$ as "column vectors" ($\mathbf{A}_{\bullet l}^T \mathbf{A}_{l\bullet} \in \mathbb{R}$, $\mathbf{A}_{\bullet l} \mathbf{A}_{l\bullet}^T \in \mathbb{R}^{D \times N}$)

Notations

- $$\mathbf{A} = \begin{bmatrix} | & | & & | \\ \mathbf{A}_{\bullet 1} & \mathbf{A}_{\bullet 2} & \dots & \mathbf{A}_{\bullet N} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} — & \mathbf{A}_{1\bullet}^T & — \\ — & \mathbf{A}_{2\bullet}^T & — \\ & \vdots & \\ — & \mathbf{A}_{D\bullet}^T & — \end{bmatrix}$$

  where $\mathbf{A} \in \mathbb{R}^{D\times N}$, $\mathbf{A}_{i\bullet} \in \mathbb{R}^N$ and $\mathbf{A}_{\bullet i} \in \mathbb{R}^D$

  Warning: We will view both columns $\mathbf{A}_{\bullet l}$ and rows $\mathbf{A}_{l\bullet}$ as "column vectors" ($\mathbf{A}_{\bullet l}^T \mathbf{A}_{l\bullet} \in \mathbb{R}$, $\mathbf{A}_{\bullet l}\mathbf{A}_{l\bullet}^T \in \mathbb{R}^{D\times N}$)

- $||\mathbf{A}||_F = \sqrt{\sum_{i=1}^{D}\sum_{j=1}^{N} A_{ij}^2}$

Notations

- $$\mathbf{A} = \begin{bmatrix} | & | & & | \\ \mathbf{A}_{\bullet 1} & \mathbf{A}_{\bullet 2} & \dots & \mathbf{A}_{\bullet N} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} — & \mathbf{A}_{1\bullet}^T & — \\ — & \mathbf{A}_{2\bullet}^T & — \\ & \vdots & \\ — & \mathbf{A}_{D\bullet}^T & — \end{bmatrix}$$

  where $\mathbf{A} \in \mathbb{R}^{D\times N}$, $\mathbf{A}_{i\bullet} \in \mathbb{R}^N$ and $\mathbf{A}_{\bullet i} \in \mathbb{R}^D$

  Warning: We will view both columns $\mathbf{A}_{\bullet l}$ and rows $\mathbf{A}_{l\bullet}$ as "column vectors" ($\mathbf{A}_{\bullet l}^T \mathbf{A}_{l\bullet} \in \mathbb{R}$, $\mathbf{A}_{\bullet l}\mathbf{A}_{l\bullet}^T \in \mathbb{R}^{D\times N}$)

- $||\mathbf{A}||_F = \sqrt{\sum_{i=1}^D \sum_{j=1}^N A_{ij}^2}$

- $||\mathbf{x}||_1 = \sum_{i=1}^D |x_i|$

Notations

- $$\mathbf{A} = \begin{bmatrix} | & | & & | \\ \mathbf{A}_{\bullet 1} & \mathbf{A}_{\bullet 2} & \dots & \mathbf{A}_{\bullet N} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} — & \mathbf{A}_{1\bullet}^T & — \\ — & \mathbf{A}_{2\bullet}^T & — \\ & \vdots & \\ — & \mathbf{A}_{D\bullet}^T & — \end{bmatrix}$$

  where $\mathbf{A} \in \mathbb{R}^{D\times N}$, $\mathbf{A}_{i\bullet} \in \mathbb{R}^N$ and $\mathbf{A}_{\bullet i} \in \mathbb{R}^D$

  Warning: We will view both columns $\mathbf{A}_{\bullet l}$ and rows $\mathbf{A}_{l\bullet}$ as "column vectors" ($\mathbf{A}_{\bullet l}^T \mathbf{A}_{l\bullet} \in \mathbb{R}$, $\mathbf{A}_{\bullet l}\mathbf{A}_{l\bullet}^T \in \mathbb{R}^{D\times N}$)

- $||\mathbf{A}||_F = \sqrt{\sum_{i=1}^D \sum_{j=1}^N A_{ij}^2}$

- $||\mathbf{x}||_1 = \sum_{i=1}^D |x_i|$

- $||\mathbf{x}||_0 = \sum_{i=1}^D \mathbb{1}_{\{x_i \neq 0\}}$

Recap of Topics: Compressed Sensing

- **Problem:** As soon as we measure a signal $\mathbf{x} \in \mathbb{R}^D$, we immediately compress and thus throw away most of the data (Think of *.raw* versus *.jpeg*)

Recap of Topics: Compressed Sensing

- **Problem:** As soon as we measure a signal $\mathbf{x} \in \mathbb{R}^D$, we immediately compress and thus throw away most of the data (Think of *.raw* versus *.jpeg*)
- **Idea:** Instead of measuring all of the signal $\mathbf{x}$, only measure linear combinations:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \in \mathbb{R}^M$$

where typically $\mathbf{W}_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \frac{1}{D})$ and $M << D$

Compressed Sensing II

- **Assume:** $\mathbf{z} = \mathbf{U}^T\mathbf{x}$ is sparse in some known fixed orthogonal basis $\mathbf{U}$

Compressed Sensing II

- **Assume:** $\mathbf{z} = \mathbf{U}^T\mathbf{x}$ is sparse in some known fixed orthogonal basis $\mathbf{U}$
- **Want to solve:**

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{U}\mathbf{z} = \boldsymbol{\Theta}\mathbf{z}$$

where $\boldsymbol{\Theta} = \mathbf{W}\mathbf{U} \in \mathbb{R}^{M \times D}$

Compressed Sensing III

- **Problem:** $\mathbf{y} = \mathbf{\Theta z}$ is an undercomplete set of equations (D unknowns versus M equations) $\implies$ Ill-posed task

Compressed Sensing III

- **Problem:** $\mathbf{y} = \mathbf{\Theta z}$ is an undercomplete set of equations (D unknowns versus M equations) $\implies$ Ill-posed task

- **Solution:** Make use of the sparsity of $\mathbf{z}$:

$$\mathbf{z}^* \in \operatorname*{argmin}_{\mathbf{z}} ||\mathbf{z}||_0 \text{ such that } \mathbf{y} = \mathbf{\Theta z}$$

  Apply the toolbox from overcomplete dictionaries: **Convex relaxation** $||\mathbf{z}||_1$ or **Matching Pursuit**

Compressed Sensing III

- **Problem:** $\mathbf{y} = \mathbf{\Theta z}$ is an undercomplete set of equations (D unknowns versus M equations) $\implies$ Ill-posed task

- **Solution:** Make use of the sparsity of $\mathbf{z}$:

$$\mathbf{z}^* \in \underset{\mathbf{z}}{\operatorname{argmin}} ||\mathbf{z}||_0 \text{ such that } \mathbf{y} = \mathbf{\Theta z}$$

  Apply the toolbox from overcomplete dictionaries: **Convex relaxation** $||\mathbf{z}||_1$ or **Matching Pursuit**

- **Notice:** Very similar problem, left-hand side consists of linear combinations $\mathbf{y}$ instead of $\mathbf{x}$.

Dictionary Learning

- **So far:** We always considered **fixed** basis $\mathbf{U}$ and found the best/sparsest linear combination $\mathbf{z}$

Dictionary Learning

- **So far:** We always considered **fixed** basis $\mathbf{U}$ and found the best/sparsest linear combination $\mathbf{z}$
- **Now:** Also **learn** the dictionary $\mathbf{U}$ $\implies$ Can adapt to characteristics of specific signal $\mathbf{x}$

Dictionary Learning

- **So far:** We always considered **fixed** basis $\mathbf{U}$ and found the best/sparsest linear combination $\mathbf{z}$
- **Now:** Also **learn** the dictionary $\mathbf{U} \implies$ Can adapt to characteristics of specific signal $\mathbf{x}$
- **Setup**: Have N samples of the signal: $\mathbf{X} \in \mathbb{R}^{D \times N}$

Dictionary Learning

- **So far:** We always considered **fixed** basis $\mathbf{U}$ and found the best/sparsest linear combination $\mathbf{z}$
- **Now:** Also **learn** the dictionary $\mathbf{U} \implies$ Can adapt to characteristics of specific signal $\mathbf{x}$
- **Setup**: Have N samples of the signal: $\mathbf{X} \in \mathbb{R}^{D \times N}$
- **Want to solve:**

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{U}\mathbf{Z}$$

where $\mathbf{U} \in \mathbb{R}^{D \times L}$ and $\mathbf{Z} \in \mathbb{R}^{L \times N}$

Dictionary Learning

- **So far:** We always considered **fixed** basis $\mathbf{U}$ and found the best/sparsest linear combination $\mathbf{z}$
- **Now:** Also **learn** the dictionary $\mathbf{U} \implies$ Can adapt to characteristics of specific signal $\mathbf{x}$
- **Setup**: Have N samples of the signal: $\mathbf{X} \in \mathbb{R}^{D \times N}$
- **Want to solve:**

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{U}\mathbf{Z}$$

where $\mathbf{U} \in \mathbb{R}^{D \times L}$ and $\mathbf{Z} \in \mathbb{R}^{L \times N}$

Matrix Factorization

- Consider the following optimization problem:

$$(\mathbf{U}^*, \mathbf{Z}^*) \in \underset{\mathbf{U},\mathbf{Z}}{\operatorname{argmin}} \, ||\mathbf{X} - \mathbf{U}\mathbf{Z}||_F^2$$

Matrix Factorization

- Consider the following optimization problem:

$$(\mathbf{U}^*, \mathbf{Z}^*) \in \underset{\mathbf{U},\mathbf{Z}}{\operatorname{argmin}} \, ||\mathbf{X} - \mathbf{U}\mathbf{Z}||_F^2$$

Recall: $||\mathbf{A}||_F^2 = \sum_{i,j=1}^{n,m} \mathbf{A}_{ij}^2$ for $\mathbf{A} \in \mathbb{R}^{m \times n}$

Matrix Factorization

- Consider the following optimization problem:

$$(\mathbf{U}^*, \mathbf{Z}^*) \in \underset{\mathbf{U},\mathbf{Z}}{\operatorname{argmin}} ||\mathbf{X} - \mathbf{U}\mathbf{Z}||_F^2$$

Recall: $||\mathbf{A}||_F^2 = \sum_{i,j=1}^{n,m} \mathbf{A}_{ij}^2$ for $\mathbf{A} \in \mathbb{R}^{m \times n}$

- **Unfortunately:** Not convex in the pair $(\mathbf{U}, \mathbf{Z})$

Matrix Factorization

- Consider the following optimization problem:

$$(\mathbf{U}^*, \mathbf{Z}^*) \in \underset{\mathbf{U},\mathbf{Z}}{\operatorname{argmin}} ||\mathbf{X} - \mathbf{UZ}||_F^2$$

Recall: $||\mathbf{A}||_F^2 = \sum_{i,j=1}^{n,m} \mathbf{A}_{ij}^2$ for $\mathbf{A} \in \mathbb{R}^{m \times n}$

- **Unfortunately:** Not convex in the pair $(\mathbf{U}, \mathbf{Z})$

- **But:** We have convexity in $\mathbf{U}$ or $\mathbf{Z}$ when one variable is kept fixed

Matrix Factorization

- Consider the following optimization problem:

$$(\mathbf{U}^*, \mathbf{Z}^*) \in \underset{\mathbf{U},\mathbf{Z}}{\operatorname{argmin}} \, ||\mathbf{X} - \mathbf{UZ}||_F^2$$

Recall: $||\mathbf{A}||_F^2 = \sum_{i,j=1}^{n,m} \mathbf{A}_{ij}^2$ for $\mathbf{A} \in \mathbb{R}^{m \times n}$

- **Unfortunately:** Not convex in the pair $(\mathbf{U}, \mathbf{Z})$

- **But:** We have convexity in $\mathbf{U}$ or $\mathbf{Z}$ when one variable is kept fixed

  $\Longrightarrow$ Step-wise **greedy** optimization

Alternating Minimization

- **Coding Step:** $\mathbf{Z}^{(t+1)} \in \text{argmin}_{\mathbf{Z}} ||\mathbf{X} - \mathbf{U}^{(t)}\mathbf{Z}||_F^2$ subject to $\mathbf{Z}$ sparse

Alternating Minimization

- **Coding Step:** $\mathbf{Z}^{(t+1)} \in \operatorname{argmin}_{\mathbf{Z}} ||\mathbf{X} - \mathbf{U}^{(t)}\mathbf{Z}||_F^2$ subject to $\mathbf{Z}$ sparse

- **Dictionary Step:** $\mathbf{U}^{(t+1)} \in \operatorname{argmin}_{\mathbf{U}} ||\mathbf{X} - \mathbf{U}\mathbf{Z}^{(t+1)}||_F^2$ subject to $||\mathbf{U}_{\bullet l}||_2 = 1$ for $l = 1, \ldots L$

Coding Step

- **Recall:** $||\mathbf{A}||_F^2 = \sum_{i=1}^n ||\mathbf{A}_{\bullet i}||_2^2$ for any $\mathbf{A} \in \mathbb{R}^{m \times n}$

Coding Step

- **Recall:** $||\mathbf{A}||_F^2 = \sum_{i=1}^n ||\mathbf{A}_{\bullet i}||_2^2$ for any $\mathbf{A} \in \mathbb{R}^{m \times n}$

- **Thus:**

$$\left\|\mathbf{X} - \mathbf{U}^{(t)}\mathbf{Z}\right\|_F^2 = \sum_{i=1}^{N} \left\|\mathbf{X}_{\bullet i} - \left(\mathbf{U}^{(t)}\mathbf{Z}\right)_{\bullet i}\right\|_2^2$$
$$= \sum_{i=1}^{N} \left\|\mathbf{X}_{\bullet i} - \mathbf{U}^{(t)}\mathbf{Z}_{\bullet i}\right\|_2^2$$

Coding Step

- **Recall:** $||\mathbf{A}||_F^2 = \sum_{i=1}^n ||\mathbf{A}_{\bullet i}||_2^2$ for any $\mathbf{A} \in \mathbb{R}^{m \times n}$

- **Thus:**

$$\begin{aligned} \left\| \mathbf{X} - \mathbf{U}^{(t)}\mathbf{Z} \right\|_F^2 &= \sum_{i=1}^{N} \left\| \mathbf{X}_{\bullet i} - \left(\mathbf{U}^{(t)}\mathbf{Z} \right)_{\bullet i} \right\|_2^2 \\ &= \sum_{i=1}^{N} \left\| \mathbf{X}_{\bullet i} - \mathbf{U}^{(t)}\mathbf{Z}_{\bullet i} \right\|_2^2 \end{aligned}$$

$\implies$ Coding step decomposes over columns of $\mathbf{Z}$

Coding Step

- **Recall:** $||\mathbf{A}||_F^2 = \sum_{i=1}^{n} ||\mathbf{A}_{\bullet i}||_2^2$ for any $\mathbf{A} \in \mathbb{R}^{m \times n}$
- **Thus:**

$$\left\| \mathbf{X} - \mathbf{U}^{(t)} \mathbf{Z} \right\|_F^2 = \sum_{i=1}^{N} \left\| \mathbf{X}_{\bullet i} - \left(\mathbf{U}^{(t)} \mathbf{Z} \right)_{\bullet i} \right\|_2^2$$
$$= \sum_{i=1}^{N} \left\| \mathbf{X}_{\bullet i} - \mathbf{U}^{(t)} \mathbf{Z}_{\bullet i} \right\|_2^2$$

$\Longrightarrow$ Coding step decomposes over columns of $\mathbf{Z}$

$$\mathbf{z}_i^{(t+1)} \in \underset{\mathbf{z}}{\operatorname{argmin}} \, ||\mathbf{z}||_0 \text{ such that } \left\| \mathbf{X}_{\bullet i} - \mathbf{U}^{(t)} \mathbf{z} \right\|_2 \leq \sigma \left\| \mathbf{X}_{\bullet i} \right\|_2$$

Dictionary Step I

- **Cannot** separate along the atoms $\mathbf{U}_{\bullet l}$

Dictionary Step I

- **Cannot** separate along the atoms $\mathbf{U}_{\bullet l}$
- Update only one atom at a time

Dictionary Step I

- **Cannot** separate along the atoms $\mathbf{U}_{\bullet l}$
- Update only one atom at a time
- Fix all atoms except for $\mathbf{U}_{\bullet l}$:

$$\mathbf{U} = \begin{bmatrix} | & & | & & | \\ \mathbf{U}_{\bullet 1}^{(t)} & \dots & \mathbf{U}_{\bullet l} & \dots & \mathbf{U}_{\bullet L}^{(t)} \\ | & & | & & | \end{bmatrix}$$

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$
- **Trick:** $\mathbf{A}\mathbf{B} = \sum_{i=1}^{d} \mathbf{A}_{\bullet i}\mathbf{B}_{i\bullet}^{T}$

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$
- **Trick:** $\mathbf{AB} = \sum_{i=1}^{d} \mathbf{A}_{\bullet i} \mathbf{B}_{i \bullet}^{T}$
- Apply the trick to the cost function:

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$
- **Trick:** $\mathbf{AB} = \sum_{i=1}^{d} \mathbf{A}_{\bullet i}\mathbf{B}_{i\bullet}^T$
- Apply the trick to the cost function:

$$\left\|\mathbf{X} - \mathbf{U}\mathbf{Z}^{(t+1)}\right\|_F^2$$

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$
- **Trick:** $\mathbf{AB} = \sum_{i=1}^{d} \mathbf{A}_{\bullet i} \mathbf{B}_{i \bullet}^{T}$
- Apply the trick to the cost function:

$$\left\| \mathbf{X} - \mathbf{U} \mathbf{Z}^{(t+1)} \right\|_F^2 = \left\| \mathbf{X} - \sum_{i=1}^{L} \mathbf{U}_{\bullet i} \left(Z_{i \bullet}^{(t+1)} \right)^T \right\|_F^2$$

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$
- **Trick:** $\mathbf{AB} = \sum_{i=1}^{d} \mathbf{A}_{\bullet i} \mathbf{B}_{i \bullet}^{T}$
- Apply the trick to the cost function:

$$\left\| \mathbf{X} - \mathbf{U}\mathbf{Z}^{(t+1)} \right\|_F^2 = \left\| \mathbf{X} - \sum_{i=1}^{L} \mathbf{U}_{\bullet i} \left(Z_{i \bullet}^{(t+1)} \right)^T \right\|_F^2$$
$$= \left\| \mathbf{X} - \sum_{i \neq l}^{L} \mathbf{U}_{\bullet i} \left(Z_{i \bullet}^{(t+1)} \right)^T - \mathbf{U}_{\bullet l} \left(\mathbf{Z}_{l \bullet}^{(t+1)} \right)^T \right\|_F^2$$

Dictionary Step II

- Isolate the residual due to $\mathbf{U}_{\bullet l}$
- **Trick:** $\mathbf{A}\mathbf{B} = \sum_{i=1}^{d} \mathbf{A}_{\bullet i}\mathbf{B}_{i\bullet}^T$
- Apply the trick to the cost function:

$$\begin{aligned}
\left\|\mathbf{X} - \mathbf{U}\mathbf{Z}^{(t+1)}\right\|_F^2 &= \left\|\mathbf{X} - \sum_{i=1}^{L} \mathbf{U}_{\bullet i}\left(Z_{i\bullet}^{(t+1)}\right)^T\right\|_F^2 \\
&= \left\|\mathbf{X} - \sum_{i\neq l}^{L} \mathbf{U}_{\bullet i}\left(Z_{i\bullet}^{(t+1)}\right)^T - \mathbf{U}_{\bullet l}\left(\mathbf{Z}_{l\bullet}^{(t+1)}\right)^T\right\|_F^2 \\
&= \left\|\mathbf{R}_l^{(t)} - \mathbf{U}_{\bullet l}\left(Z_{l\bullet}^{(t+1)}\right)^T\right\|_F^2
\end{aligned}$$

Dictionary Step III

- **Recall:** $\mathbf{u}\mathbf{v}^T \in \mathbb{R}^{m \times n}$ for $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$ is a rank 1 matrix

Dictionary Step III

- **Recall:** $\mathbf{u}\mathbf{v}^T \in \mathbb{R}^{m \times n}$ for $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$ is a rank 1 matrix
- **Thus:** We're making a rank 1 approximation to $\mathbf{R}_l^{(t)}$

Dictionary Step III

- **Recall:** $\mathbf{u}\mathbf{v}^T \in \mathbb{R}^{m \times n}$ for $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$ is a rank 1 matrix
- **Thus:** We're making a rank 1 approximation to $\mathbf{R}_l^{(t)}$
- The best solution is given by the SVD:

Dictionary Step III

- **Recall:** $\mathbf{u}\mathbf{v}^T \in \mathbb{R}^{m \times n}$ for $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$ is a rank 1 matrix
- **Thus:** We're making a rank 1 approximation to $\mathbf{R}_l^{(t)}$
- The best solution is given by the SVD:

$$\mathbf{R}_l^{(t)} = \tilde{\mathbf{U}}\mathbf{\Sigma}\tilde{\mathbf{V}}^T = \sum_i \sigma_i \tilde{\mathbf{U}}_{\bullet i} \tilde{\mathbf{V}}_{i\bullet}^T$$

Dictionary Step III

- **Recall:** $\mathbf{u}\mathbf{v}^T \in \mathbb{R}^{m \times n}$ for $\mathbf{u} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^n$ is a rank 1 matrix
- **Thus:** We're making a rank 1 approximation to $\mathbf{R}_l^{(t)}$
- The best solution is given by the SVD:

$$\mathbf{R}_l^{(t)} = \tilde{\mathbf{U}}\mathbf{\Sigma}\tilde{\mathbf{V}}^T = \sum_i \sigma_i \tilde{\mathbf{U}}_{\bullet i} \tilde{\mathbf{V}}_{i\bullet}^T$$

$\implies \mathbf{U}_{\bullet l}^* = \tilde{\mathbf{U}}_{\bullet 1}$ and $\|\mathbf{U}_{\bullet l}^*\|_2 = 1$ "for free"

Exercise 1a)

- Given signal $\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ and $\mathbf{U} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$

Exercise 1a)

- Given signal $\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ and $\mathbf{U} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$
- **Claim 1:** $||\mathbf{x} - z\mathbf{u}||_2$ is minimal iff $|\mathbf{x}^T\mathbf{u}|$ is maximal:

Exercise 1a)

- Given signal $\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ and $\mathbf{U} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$

- **Claim 1:** $||\mathbf{x} - z\mathbf{u}||_2$ is minimal iff $|\mathbf{x}^T\mathbf{u}|$ is maximal:

  Indeed:

$$\begin{aligned} \|\mathbf{x} - z\mathbf{u}\|_2^2 &= \|\mathbf{x}\|_2^2 + z^2 \|\mathbf{u}\|_2^2 - 2z\mathbf{x}^T\mathbf{u} \\ &= \|\mathbf{x}\|_2^2 + z^2 - 2z\mathbf{x}^T\mathbf{u} \end{aligned}$$

Exercise 1a)

- Given signal $\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ and $\mathbf{U} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$

- **Claim 1:** $||\mathbf{x} - z\mathbf{u}||_2$ is minimal iff $|\mathbf{x}^T\mathbf{u}|$ is maximal:

  Indeed:

$$\begin{aligned} \|\mathbf{x} - z\mathbf{u}\|_2^2 &= \|\mathbf{x}\|_2^2 + z^2 \|\mathbf{u}\|_2^2 - 2z\mathbf{x}^T\mathbf{u} \\ &= \|\mathbf{x}\|_2^2 + z^2 - 2z\mathbf{x}^T\mathbf{u} \end{aligned}$$

- **Claim 2:** $z^* = \mathbf{x}^T\mathbf{u}$

Exercise 1a)

- Given signal $\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ and $\mathbf{U} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$

- **Claim 1:** $||\mathbf{x} - z\mathbf{u}||_2$ is minimal iff $|\mathbf{x}^T\mathbf{u}|$ is maximal:

  Indeed:

$$\begin{aligned} \|\mathbf{x} - z\mathbf{u}\|_2^2 &= \|\mathbf{x}\|_2^2 + z^2 \|\mathbf{u}\|_2^2 - 2z\mathbf{x}^T\mathbf{u} \\ &= \|\mathbf{x}\|_2^2 + z^2 - 2z\mathbf{x}^T\mathbf{u} \end{aligned}$$

- **Claim 2:** $z^* = \mathbf{x}^T\mathbf{u}$

  Indeed:

$$\frac{\partial}{\partial z}||\mathbf{x} - z\mathbf{u}||_2^2 \stackrel{!}{=} 0 \iff 2z - 2\mathbf{x}^T\mathbf{u} \stackrel{!}{=} 0 \iff z = \mathbf{x}^T\mathbf{u}$$

Exercise 1a)

- Using Claim 1, we only have to calculate the inner products:

$$|\mathbf{x}^T \mathbf{U}_{\bullet 1}| = \frac{2}{\sqrt{3}},\ |\mathbf{x}^T \mathbf{U}_{\bullet 2}| = \frac{4}{\sqrt{3}},\ |\mathbf{x}^T \mathbf{U}_{\bullet 3}| = 0,\ |\mathbf{x}^T \mathbf{U}_{\bullet 4}| = \frac{6}{\sqrt{3}}$$

Exercise 1a)

- Using Claim 1, we only have to calculate the inner products:

$$|\mathbf{x}^T \mathbf{U}_{\bullet 1}| = \frac{2}{\sqrt{3}},\ |\mathbf{x}^T \mathbf{U}_{\bullet 2}| = \frac{4}{\sqrt{3}},\ |\mathbf{x}^T \mathbf{U}_{\bullet 3}| = 0,\ |\mathbf{x}^T \mathbf{U}_{\bullet 4}| = \frac{6}{\sqrt{3}}$$

Exercise 1a)

- Using Claim 1, we only have to calculate the inner products:

$$|\mathbf{x}^T \mathbf{U}_{\bullet 1}| = \frac{2}{\sqrt{3}},\ |\mathbf{x}^T \mathbf{U}_{\bullet 2}| = \frac{4}{\sqrt{3}},\ |\mathbf{x}^T \mathbf{U}_{\bullet 3}| = 0,\ |\mathbf{x}^T \mathbf{U}_{\bullet 4}| = \frac{6}{\sqrt{3}}$$

- $\hat{\mathbf{x}}^{(1)} = \left(\mathbf{x}^T \mathbf{U}_{\bullet 4}\right) \mathbf{U}_{\bullet 4} = \begin{pmatrix} 2 \\ 2 \\ -2 \end{pmatrix}$

Exercise 1a)

- Using Claim 1, we only have to calculate the inner products:

$$|\mathbf{x}^T\mathbf{U}_{\bullet 1}| = \frac{2}{\sqrt{3}},\ |\mathbf{x}^T\mathbf{U}_{\bullet 2}| = \frac{4}{\sqrt{3}},\ |\mathbf{x}^T\mathbf{U}_{\bullet 3}| = 0,\ |\mathbf{x}^T\mathbf{U}_{\bullet 4}| = \frac{6}{\sqrt{3}}$$

- $\hat{\mathbf{x}}^{(1)} = \left(\mathbf{x}^T\mathbf{U}_{\bullet 4}\right)\mathbf{U}_{\bullet 4} = \begin{pmatrix} 2 \\ 2 \\ -2 \end{pmatrix}$

- $\mathbf{r}^{(1)} = \mathbf{x} - \hat{\mathbf{x}}^{(0)} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

Exercise 1b)

- **Now:** Find $||\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(2)}||_2 = ||\mathbf{r}^{(1)} - z\mathbf{u}^{(2)}||$

Exercise 1b)

- **Now:** Find $||\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(2)}||_2 = ||\mathbf{r}^{(1)} - z\mathbf{u}^{(2)}||$
- We can again calculate the inner products:

  $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = 0$, $|\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{2}{\sqrt{3}}$, $|\mathbf{U}_{\bullet 4}^T \mathbf{r}^{(1)}| = 0$

Exercise 1b)

- **Now:** Find $||\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(2)}||_2 = ||\mathbf{r}^{(1)} - z\mathbf{u}^{(2)}||$
- We can again calculate the inner products:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = 0, \; |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{2}{\sqrt{3}}, \; |\mathbf{U}_{\bullet 4}^T \mathbf{r}^{(1)}| = 0$$

Exercise 1b)

- **Now:** Find $||\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(2)}||_2 = ||\mathbf{r}^{(1)} - z\mathbf{u}^{(2)}||$
- We can again calculate the inner products:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = 0, \; |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{2}{\sqrt{3}}, \; |\mathbf{U}_{\bullet 4}^T \mathbf{r}^{(1)}| = 0$$

- We can take either $\mathbf{U}_{\bullet,2}$ or $\mathbf{U}_{\bullet,3}$, let's go with $\mathbf{U}_{\bullet,2}$

Exercise 1b)

- **Now:** Find $||\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(2)}||_2 = ||\mathbf{r}^{(1)} - z\mathbf{u}^{(2)}||$
- We can again calculate the inner products:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = 0, \; |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{2}{\sqrt{3}}, \; |\mathbf{U}_{\bullet 4}^T \mathbf{r}^{(1)}| = 0$$

- We can take either $\mathbf{U}_{\bullet,2}$ or $\mathbf{U}_{\bullet,3}$, let's go with $\mathbf{U}_{\bullet,2}$
- $\hat{\mathbf{x}}^{(2)} = \left(\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 2} = \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ -\frac{2}{3} \end{pmatrix}$

Exercise 1c)

We can now obtain the sparse representations:

Exercise 1c)

We can now obtain the sparse representations:

- $||\mathbf{z}||_0 = 1 \implies \mathbf{z} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{6}{\sqrt{3}} \end{pmatrix}$

Exercise 1c)

We can now obtain the sparse representations:

- $||\mathbf{z}||_0 = 1 \implies \mathbf{z} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{6}{\sqrt{3}} \end{pmatrix}$

- $||\mathbf{z}||_0 = 2 \implies \mathbf{z} = \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{3}} \\ 0 \\ \frac{6}{\sqrt{3}} \end{pmatrix}$ or $\mathbf{z} = \begin{pmatrix} 0 \\ 0 \\ \frac{2}{\sqrt{3}} \\ \frac{6}{\sqrt{3}} \end{pmatrix}$

Exercise 2.a)

Draw the level set $\{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x}||_i = 1\}$ for $i \in \{0, 1, 2\}$

Exercise 2.a)

Draw the level set $\{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x}||_i = 1\}$ for $i \in \{0, 1, 2\}$

- $||\mathbf{x}||_2 = 1 \iff x_1^2 + x_2^2 = 1 \implies$ Circle with center $(0, 0)$ and radius $1 \implies$ Picture b)

Exercise 2.a)

Draw the level set $\{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x}||_i = 1\}$ for $i \in \{0, 1, 2\}$

- $||\mathbf{x}||_2 = 1 \iff x_1^2 + x_2^2 = 1 \implies$ Circle with center $(0, 0)$ and radius $1 \implies$ Picture b)

- $||\mathbf{x}||_1 = 1 \iff |x_1| + |x_2| = 1$

  Let's look at 1st quadrant: $x_1, x_2 > 0 \implies x_1 + x_2 = 1$

  Equation of a line with slope -1 and intercept 1.

  Similarly for other quadrants $\implies$ Picture a)

Exercise 2.a)

Draw the level set $\{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x}||_i = 1\}$ for $i \in \{0, 1, 2\}$

- $||\mathbf{x}||_2 = 1 \iff x_1^2 + x_2^2 = 1 \implies$ Circle with center $(0, 0)$ and radius $1 \implies$ Picture b)

- $||\mathbf{x}||_1 = 1 \iff |x_1| + |x_2| = 1$

  Let's look at 1st quadrant: $x_1, x_2 > 0 \implies x_1 + x_2 = 1$

  Equation of a line with slope -1 and intercept 1.

  Similarly for other quadrants $\implies$ Picture a)

- $||\mathbf{x}||_0 = 1 \iff \mathbb{I}_{\{x_1 \neq 0\}} + \mathbb{I}_{\{x_2 \neq 0\}} = 1$

  $\iff (a, 0)$ or $(0, a)$ for $a \in \mathbb{R} \implies$ Picture c)

Exercise 2.b)

Goal: $\min_{\mathbf{x}\in\mathbb{R}^2} ||\mathbf{x}||_i$ such that $\frac{1}{2}x_1 + x_2 = 1$

Intuition: Draw the line and $||\mathbf{x}||_i = r$ for smaller and smaller r until there is no intersection. Remember, we need an intersection, otherwise we cannot fulfill the constraint.

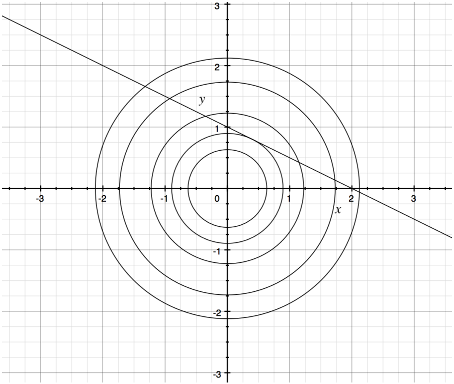

Figure: Illustration for $||\cdot||_2$

Exercise 2.b)

The same procedure leads to the following plots:

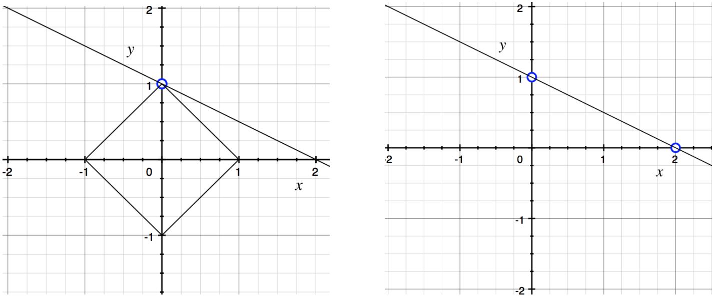

Figure: Solution for $||\cdot||_1$ and $||\cdot||_0$ respectively

Exercise 2.c)

- We can see that $||\cdot||_1$ and $||\cdot||_0$ share the solution $\mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Exercise 2.c)

- We can see that $||\cdot||_1$ and $||\cdot||_0$ share the solution $\mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$
- This gives some (limited) intuition why we can often relax an optimization problem involving $||\cdot||_0$ to $||\cdot||_1$ **without** affecting the set of solutions too much.

Exercise 4a)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that $\hat{\mathbf{x}}$ will never converge to $\mathbf{x}$

Exercise 4a)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that $\hat{\mathbf{x}}$ will never converge to $\mathbf{x}$
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{pmatrix}$

Exercise 4a)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that $\hat{\mathbf{x}}$ will never converge to $\mathbf{x}$
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{pmatrix}$
- First consider signals of the form $\mathbf{x} = \begin{pmatrix} 0 \\ z \end{pmatrix}$ for $z \in \mathbb{R}$ and let's perform the first step of matching pursuit:

  $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 0 \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = |z|\frac{\sqrt{2}}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{|z|}{2}$

Exercise 4a)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that $\hat{\mathbf{x}}$ will never converge to $\mathbf{x}$
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{pmatrix}$
- First consider signals of the form $\mathbf{x} = \begin{pmatrix} 0 \\ z \end{pmatrix}$ for $z \in \mathbb{R}$ and let's perform the first step of matching pursuit:

  $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 0 \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = |z| \frac{\sqrt{2}}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{|z|}{2}$

Exercise 4a)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that $\hat{\mathbf{x}}$ will never converge to $\mathbf{x}$
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{pmatrix}$
- First consider signals of the form $\mathbf{x} = \begin{pmatrix} 0 \\ z \end{pmatrix}$ for $z \in \mathbb{R}$ and let's perform the first step of matching pursuit:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 0 \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = |z|\frac{\sqrt{2}}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{|z|}{2}$$

$$\implies \mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 2}\right) \mathbf{U}_{\bullet 2} = \begin{pmatrix} -\frac{z}{2} \\ \frac{z}{2} \end{pmatrix} = \begin{pmatrix} -y \\ y \end{pmatrix}$$

Exercise 4a) Continued

- Let us perform the next step:

Exercise 4a) Continued

- Let us perform the next step:

  $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$

Exercise 4a) Continued

- Let us perform the next step:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$$

Exercise 4a) Continued

- Let us perform the next step:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$$

$$\implies \mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{z}{2} \end{pmatrix}$$

Exercise 4a) Continued

- Let us perform the next step:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$$

$$\implies \mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{z}{2} \end{pmatrix}$$

- We're again in the same setting as before!

Exercise 4a) Continued

- Let us perform the next step:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$$

$$\implies \mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{z}{2} \end{pmatrix}$$

- We're again in the same setting as before!

- **Induction:** $\mathbf{r}^{(2n)} = \begin{pmatrix} 0 \\ \frac{z}{2^n} \end{pmatrix}$ and $\mathbf{r}^{(2n+1)} = \begin{pmatrix} -\frac{z}{2^{n+1}} \\ \frac{z}{2^{n+1}} \end{pmatrix}$

Exercise 4a) Continued

- Let us perform the next step:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$$

$$\implies \mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{z}{2} \end{pmatrix}$$

- We're again in the same setting as before!

- **Induction:** $\mathbf{r}^{(2n)} = \begin{pmatrix} 0 \\ \frac{z}{2^n} \end{pmatrix}$ and $\mathbf{r}^{(2n+1)} = \begin{pmatrix} -\frac{z}{2^{n+1}} \\ \frac{z}{2^{n+1}} \end{pmatrix}$

  Only for $n \to \infty$ we can reach $\mathbf{r}^{(\infty)} = \mathbf{0}$, e.g. for $\mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Exercise 4a) Continued

- Let us perform the next step:

$$|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = |y|, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = 0, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = \frac{\sqrt{3}-1}{4}|y|$$

$$\implies \mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{z}{2} \end{pmatrix}$$

- We're again in the same setting as before!

- **Induction:** $\mathbf{r}^{(2n)} = \begin{pmatrix} 0 \\ \frac{z}{2^n} \end{pmatrix}$ and $\mathbf{r}^{(2n+1)} = \begin{pmatrix} -\frac{z}{2^{n+1}} \\ \frac{z}{2^{n+1}} \end{pmatrix}$

  Only for $n \to \infty$ we can reach $\mathbf{r}^{(\infty)} = \mathbf{0}$, e.g. for $\mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

- **So:** MP converges, but it takes ∞ steps!

Exercise 4.b)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that the sparse representation found by MP is not optimal

Exercise 4.b)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that the sparse representation found by MP is not optimal
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

Exercise 4.b)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that the sparse representation found by MP is not optimal
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$
- **Notice:** $\mathbf{x} = 2 \cdot \mathbf{U}_{\bullet 1} + 1 \cdot \mathbf{U}_{\bullet 2} + 0 \cdot \mathbf{U}_{\bullet 3} \implies \mathbf{z} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$

Exercise 4.b)

- **Goal:** Find a overcomplete dictionary $\mathbf{U}$ and signal $\mathbf{x}$ such that the sparse representation found by MP is not optimal
- Let's choose $\mathbf{U} = \begin{pmatrix} 1 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & \frac{\sqrt{2}}{2} \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$
- **Notice:** $\mathbf{x} = 2 \cdot \mathbf{U}_{\bullet 1} + 1 \cdot \mathbf{U}_{\bullet 2} + 0 \cdot \mathbf{U}_{\bullet 3} \implies \mathbf{z} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$
- We found a representation $\mathbf{z}$ with $||\mathbf{z}||_0 = 2$

Exercise 4.b) Continued

What does Matching Pursuit return?

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

2a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = 0$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

2a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = 0$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

2a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = 0$

2b $\mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 3} = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

2a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = 0$

2b $\mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 3} = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$

3a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(2)}| = 0, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(2)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(2)}| = \frac{\sqrt{2}}{4}$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

2a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = 0$

2b $\mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 3} = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$

3a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(2)}| = 0, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(2)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(2)}| = \frac{\sqrt{2}}{4}$

Exercise 4.b) Continued

What does Matching Pursuit return?

1a $|\mathbf{U}_{\bullet 1}^T \mathbf{x}| = 2, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{x}| = 1, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{x}| = \frac{3}{\sqrt{2}}$

1b $\mathbf{r}^{(1)} = \mathbf{x} - \left(\mathbf{x}^T \mathbf{U}_{\bullet 1}\right) \mathbf{U}_{\bullet 1} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

2a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(1)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}| = 0$

2b $\mathbf{r}^{(2)} = \mathbf{r}^{(1)} - \left(\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(1)}\right) \mathbf{U}_{\bullet 3} = \begin{pmatrix} 0 \\ -\frac{1}{2} \end{pmatrix}$

3a $|\mathbf{U}_{\bullet 1}^T \mathbf{r}^{(2)}| = 0, \quad |\mathbf{U}_{\bullet 2}^T \mathbf{r}^{(2)}| = \frac{1}{2}, \quad |\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(2)}| = \frac{\sqrt{2}}{4}$

3b $\mathbf{r}^{(3)} = \mathbf{r}^{(2)} - \left(\mathbf{U}_{\bullet 3}^T \mathbf{r}^{(2)}\right) \mathbf{U}_{\bullet 3} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Exercise 4.b) Continued

- Matching Pursuit converges ($\mathbf{r}^{(3)} = 0$) after 3 steps

Exercise 4.b) Continued

- Matching Pursuit converges ($\mathbf{r}^{(3)} = 0$) after 3 steps
- We obtain the representation $\mathbf{z}_{MP} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{\sqrt{2}} \end{pmatrix}$

Exercise 4.b) Continued

- Matching Pursuit converges ($\mathbf{r}^{(3)} = 0$) after 3 steps
- We obtain the representation $\mathbf{z}_{MP} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{\sqrt{2}} \end{pmatrix}$
- Notice that $||\mathbf{z}_{MP}||_0 = 3 > ||\mathbf{z}||_0 = 2$

Exercise 4.b) Continued

- Matching Pursuit converges ($\mathbf{r}^{(3)} = 0$) after 3 steps
- We obtain the representation $\mathbf{z}_{MP} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{\sqrt{2}} \end{pmatrix}$
- Notice that $||\mathbf{z}_{MP}||_0 = 3 > ||\mathbf{z}||_0 = 2$

$\implies$ Matching Pursuit didn't find the sparsest representation