Dictionary Learning & Compressed Sensing

Antonio Orvieto, Simon Meinhard (based on slides from Francesco Locatello)

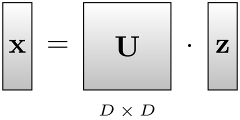
May 16, 2019

Overview

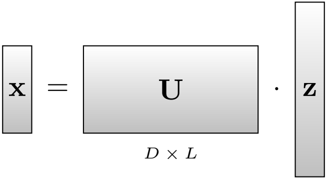
- Review: Overcomplete dictionaries
- Matching pursuit (continuation)
- Dictionary learning
- Exercises

Review: Overcomplete dictionaries

Sparse coding with a complete dictionary:



Sparse coding with an overcomplete dictionary ($L > D$):



Why overcomplete? 2D example

Figure: Original Image

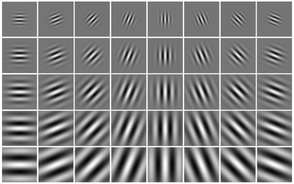


Figure: Gabor Basis

Why overcomplete? 2D example

Figure: Original Image

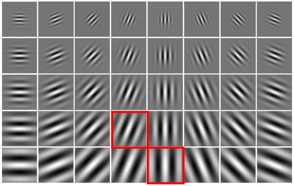


Figure: Gabor Basis

Signal Reconstruction (General Dictionary)

$\mathbf{U} \in \mathbb{R}^{D \times L}$ is overcomplete $(L > D)$:

- Ill-posed problem: more unknowns than equations, encoding not unique!
- add constraint: find sparsest $\mathbf{z} \in \mathbb{R}^L$ such that $\mathbf{x} = \mathbf{U}\mathbf{z}$

Solve mathematical program

$$\begin{aligned} \mathbf{z}^\star \quad &\in \quad \arg\min_{\mathbf{z}} \|\mathbf{z}\|_0 \\ \text{s.t.} \quad &\quad\quad \mathbf{x} = \mathbf{U}\mathbf{z} \end{aligned}$$

- $\|\mathbf{z}\|_0$ counts the number of non-zero elements in $\mathbf{z}$.

Roadmap to Solution

Original Problem is NP-Hard: How to Proceed?

1. Use a greedy approximation algorithm (Matching Pursuit)

Roadmap to Solution

Original Problem is NP-Hard: How to Proceed?

1. Use a greedy approximation algorithm (Matching Pursuit)
2. Relax the $\|\mathbf{z}\|_0$ "norm" so that it becomes a convex optimization problem

Roadmap to Solution

Original Problem is NP-Hard: How to Proceed?

1. Use a greedy approximation algorithm (Matching Pursuit)
2. Relax the $\|\mathbf{z}\|_0$ "norm" so that it becomes a convex optimization problem

MP Algorithm

1: $\mathbf{z} \leftarrow \mathbf{0}$, $\mathbf{r} \leftarrow \mathbf{x}$

2: **while** Approximation not satisfactory **do**

3: Select atom with maximum absolute correlation to residual:

$$d^* \leftarrow \underset{d}{\operatorname{argmax}} \, |\langle \mathbf{u}_d, \mathbf{r} \rangle|$$

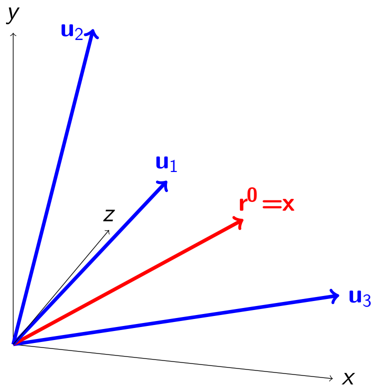
4: Update coefficient vector and residual:

$$z_{d^*} \leftarrow z_{d^*} + \langle \mathbf{u}_{d^*}, \mathbf{r} \rangle$$

$$\mathbf{r} \leftarrow \mathbf{r} - \langle \mathbf{u}_{d^*}, \mathbf{r} \rangle \mathbf{u}_{d^*}$$

5: **end while**

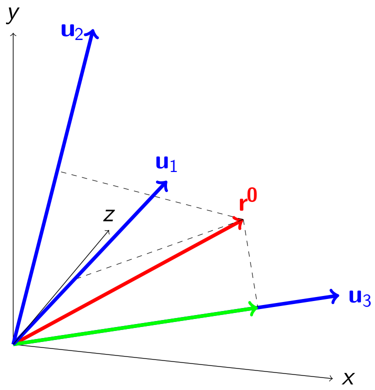
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0, 0)^\top$$

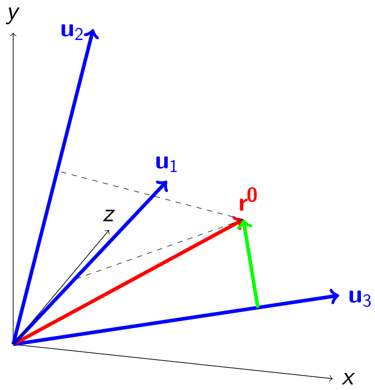
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0, 0)^\top$$

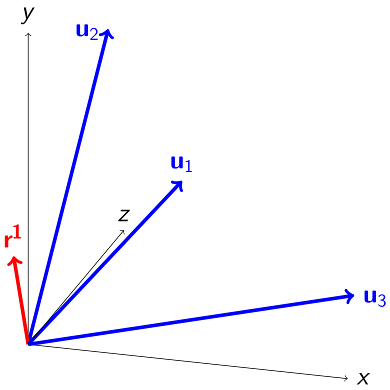
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0, 0.75)^\top$$

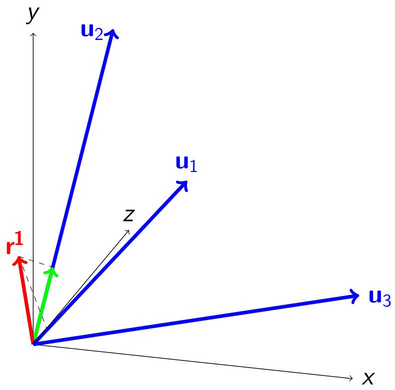
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0, 0.75)^\top$$

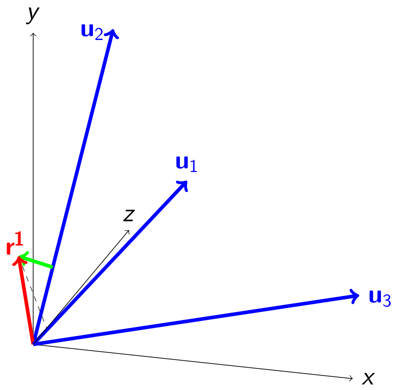
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0, 0.75)^\top$$

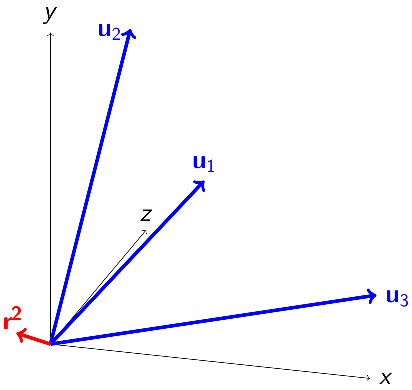
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0.24, 0.75)^\top$$

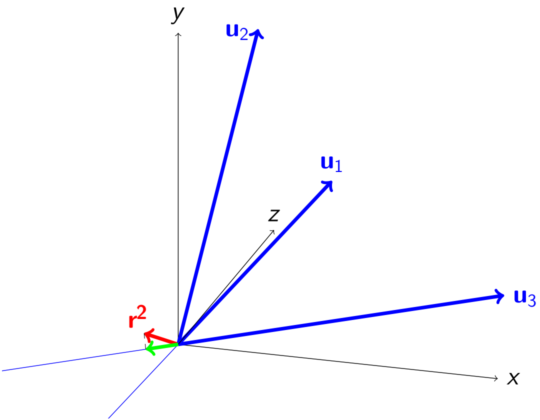
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0.24, 0.75)^\top$$

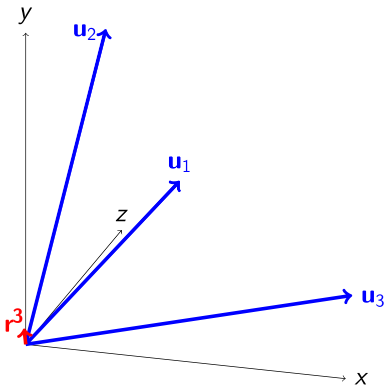
MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0.24, 0.65)^\top$$

MP Example



Bach et al. (2009)

$$\mathbf{z} = (0, 0.24, 0.65)^\top$$

Matching Pursuit - Properties

Last week:

- Algorithm converges to $\mathbf{x}$
- Convergence speed

This week:

- Algorithm is greedy, reduces residual energy as much as possible for every iteration
- Orthogonality between atoms and residual

Matching Pursuit - Minimizing the Residual

Matching pursuit greedily reduces the residual energy at every iteration.

Proof for first iteration:

- Project $\mathbf{r}^0 = \mathbf{x}$ on atom $\mathbf{u}_d$, to get

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{u}_d \rangle \, \mathbf{u}_d + \mathbf{r}^1$$

- (Seen last week): Since $\mathbf{r}^1$ is orthogonal to $\mathbf{u}_d$, and $\|\mathbf{u}_d\|_2 = 1$,

$$\|\mathbf{x}\|_2^2 = |\langle \mathbf{x}, \mathbf{u}_d \rangle|^2 + \left\|\mathbf{r}^1\right\|_2^2$$

- Therefore, $\left\|\mathbf{r}^1\right\|_2^2$ is minimized by maximizing $\left|\left\langle \mathbf{r}^0, \mathbf{u}_d \right\rangle\right|^2$.

Orthogonality Between Atoms and Residual

$$d^*(t) := \underset{d}{\operatorname{argmax}} \left| \left\langle \mathbf{r}^t, \mathbf{u}_d \right\rangle \right|$$

Orthogonality Between Atoms and Residual

$$d^*(t) := \underset{d}{\operatorname{argmax}} \left| \left\langle \mathbf{r}^t, \mathbf{u}_d \right\rangle \right|$$

From the algorithm, we have the following:

$$\mathbf{x} = \left\langle \mathbf{x}, \mathbf{u}_{d^*(1)} \right\rangle \mathbf{u}_{d^*(1)} + \mathbf{r}^1$$

$\mathbf{r}^1$ is orthogonal to $\mathbf{u}_{d^*(1)}$

Orthogonality Between Atoms and Residual

$$d^*(t) := \underset{d}{\operatorname{argmax}} \left|\left\langle \mathbf{r}^t, \mathbf{u}_d \right\rangle\right|$$

From the algorithm, we have the following:

$$\mathbf{x} = \left\langle \mathbf{x}, \mathbf{u}_{d^*(1)} \right\rangle \mathbf{u}_{d^*(1)} + \mathbf{r}^1$$

$\mathbf{r}^1$ is orthogonal to $\mathbf{u}_{d^*(1)}$

For the next step we have

$$\mathbf{r}^1 = \left\langle \mathbf{r}^1, \mathbf{u}_{d^*(2)} \right\rangle \mathbf{u}_{d^*(2)} + \mathbf{r}^2$$

$\mathbf{r}^2$ is orthogonal to $\mathbf{u}_{d^*(2)}$

Orthogonality Between Atoms and Residual

$$d^*(t) := \underset{d}{\operatorname{argmax}} \left| \left\langle \mathbf{r}^t, \mathbf{u}_d \right\rangle \right|$$

From the algorithm, we have the following:

$$\mathbf{x} = \left\langle \mathbf{x}, \mathbf{u}_{d^*(1)} \right\rangle \mathbf{u}_{d^*(1)} + \mathbf{r}^1$$

$\mathbf{r}^1$ is orthogonal to $\mathbf{u}_{d^*(1)}$

For the next step we have

$$\mathbf{r}^1 = \left\langle \mathbf{r}^1, \mathbf{u}_{d^*(2)} \right\rangle \mathbf{u}_{d^*(2)} + \mathbf{r}^2$$

$\mathbf{r}^2$ is orthogonal to $\mathbf{u}_{d^*(2)}$

Question: Is $\mathbf{r}^2$ orthogonal to $\mathbf{u}_{d^*(1)}$? When is it true?

Orthogonality Between Atoms and Residual

$$d^*(t) := \underset{d}{\operatorname{argmax}} \left| \left\langle \mathbf{r}^t, \mathbf{u}_d \right\rangle \right|$$

From the algorithm, we have the following:

$$\mathbf{x} = \left\langle \mathbf{x}, \mathbf{u}_{d^*(1)} \right\rangle \mathbf{u}_{d^*(1)} + \mathbf{r}^1$$

$\mathbf{r}^1$ is orthogonal to $\mathbf{u}_{d^*(1)}$

For the next step we have

$$\mathbf{r}^1 = \left\langle \mathbf{r}^1, \mathbf{u}_{d^*(2)} \right\rangle \mathbf{u}_{d^*(2)} + \mathbf{r}^2$$

$\mathbf{r}^2$ is orthogonal to $\mathbf{u}_{d^*(2)}$

Question: Is $\mathbf{r}^2$ orthogonal to $\mathbf{u}_{d^*(1)}$? When is it true?
Solution:

$$\langle \mathbf{r}^2, \mathbf{u}_{d(1)} \rangle = -\langle \mathbf{r}^1, \mathbf{u}_{d(2)} \rangle \langle \mathbf{u}_{d(2)}, \mathbf{u}_{d(1)} \rangle$$

Orthogonality Between Atoms and Residual

$$d^*(t) := \underset{d}{\operatorname{argmax}} \left| \left\langle \mathbf{r}^t, \mathbf{u}_d \right\rangle \right|$$

From the algorithm, we have the following:

$$\mathbf{x} = \left\langle \mathbf{x}, \mathbf{u}_{d^*(1)} \right\rangle \mathbf{u}_{d^*(1)} + \mathbf{r}^1$$

$\mathbf{r}^1$ is orthogonal to $\mathbf{u}_{d^*(1)}$

For the next step we have

$$\mathbf{r}^1 = \left\langle \mathbf{r}^1, \mathbf{u}_{d^*(2)} \right\rangle \mathbf{u}_{d^*(2)} + \mathbf{r}^2$$

$\mathbf{r}^2$ is orthogonal to $\mathbf{u}_{d^*(2)}$

Question: Is $\mathbf{r}^2$ orthogonal to $\mathbf{u}_{d^*(1)}$? When is it true?
Solution:

$$\langle \mathbf{r}^2, \mathbf{u}_{d(1)} \rangle = -\langle \mathbf{r}^1, \mathbf{u}_{d(2)} \rangle \langle \mathbf{u}_{d(2)}, \mathbf{u}_{d(1)} \rangle$$

What happens if we have an orthonormal dictionary?

Sparse Coding, Learning the Dictionary

Previously, we had the overcomplete dictionary $\mathbf{U}$ fixed and we want to compute sparse representation $\mathbf{z}$ of a single vector $\mathbf{x}$:

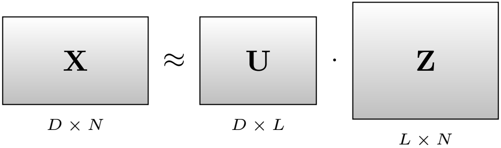
$$\mathbf{x} = \underbrace{\mathbf{U}}_{D \times L} \cdot \mathbf{z}$$

Now, we have several vectors $\mathbf{x}_1, \cdots, \mathbf{x}_N$ for which we want to find a suitable overcomplete dictionary and sparse representations $\mathbf{z}_1, \cdots, \mathbf{z}_N$:

$$\underbrace{\mathbf{X}}_{D \times N} \approx \underbrace{\mathbf{U}}_{D \times L} \cdot \underbrace{\mathbf{Z}}_{L \times N}$$

Sparse Coding, Learning the Dictionary

we have several vectors $\mathbf{x}_1, \cdots, \mathbf{x}_N$ for which we want to find a suitable overcomplete dictionary and sparse representations $\mathbf{z}_1, \cdots, \mathbf{z}_N$:



For data matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ consisting of N vectors in $\mathbb{R}^D$, want to find:

- Overcomplete dictionary $\mathbf{U}^* \in \mathbb{R}^{D \times L}$ with L unit normed atoms
- Sparse representations $\mathbf{Z}^* \in \mathbb{R}^{L \times N}$ of vectors in data matrix $\mathbf{X}$

such that $\|\mathbf{X} - \mathbf{U}^*\mathbf{Z}^*\|_F^2$ is minimized.

High-level Algorithm

Find

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

over all suitable dictionaries and sparse representations.

High-level Algorithm

Find

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

over all suitable dictionaries and sparse representations.

- Objective function not convex! :-(

High-level Algorithm

Find

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

over all suitable dictionaries and sparse representations.

- Objective function not convex! :-(

Iterative greedy minimization

repeat:

High-level Algorithm

Find

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

over all suitable dictionaries and sparse representations.

- Objective function not convex! :-(

Iterative greedy minimization

repeat:

1. **Coding step** Improve sparse representations $\mathbf{Z}$, i.e. $\mathbf{Z}^{t+1} \in \arg\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{U}^t\mathbf{Z}\|_F^2$ while keeping $\mathbf{Z}$ sparse and $\mathbf{U}$ fixed.

High-level Algorithm

Find

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

over all suitable dictionaries and sparse representations.

- Objective function not convex! :-(

Iterative greedy minimization

repeat:

1. **Coding step** Improve sparse representations $\mathbf{Z}$, i.e. $\mathbf{Z}^{t+1} \in \arg\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{U}^t\mathbf{Z}\|_F^2$ while keeping $\mathbf{Z}$ sparse and $\mathbf{U}$ fixed.
2. **Dictionary update step**: Improve dictionary $\mathbf{U}$, i.e. $\mathbf{U}^{t+1} \in \arg\min_{\mathbf{U}} \|\mathbf{X} - \mathbf{U}\mathbf{Z}^{t+1}\|_F^2$ while maintaining the normalization of the atoms and $\mathbf{Z}$ "fixed".

Coding Step

Improve sparse representations

$$\mathbf{Z}^{t+1} \in \underset{\mathbf{Z}}{\operatorname{argmin}} \left\| \mathbf{X} - \mathbf{U}^t \mathbf{Z} \right\|_F^2$$

- Not convex, why?
- Intuitively: $\mathbf{x}_i$ depends only on $\mathbf{z}_i$ (and no other $\mathbf{z}_j$) and $\mathbf{U}$.
- Squared Frobenius norm $\implies$ independently find best sparse representation for every vector in $\mathbf{X}$.

Dictionary Update I

Improve dictionary

$$\mathbf{U}^{t+1} \in \underset{\mathbf{U}}{\operatorname{argmin}} \left\| \mathbf{X} - \mathbf{U}\mathbf{Z}^{t+1} \right\|_F^2$$

- Residual not separable in atoms (columns of $\mathbf{U}$)
- Approximation: update one atom at a time (loop over all atoms $\mathbf{u}_l$)
    1. Set $\mathbf{U} = [\mathbf{u}_1^t \cdots \mathbf{u}_l \cdots \mathbf{u}_L^t]$, i.e. fix all atoms except $\mathbf{u}_l$.
    2. Isolate $\mathbf{R}_l^t$, the residual that is due to atom $\mathbf{u}_l$.
    3. Find $\mathbf{u}_l^*$ that minimizes $\mathbf{R}_l^t$, subject to $\|\mathbf{u}_l^*\|_2 = 1$.

Dictionary Update II

- Isolate $\mathbf{R}_l^t$: residual due to atom $\mathbf{u}_l$

$$
\begin{aligned}
& \left\| \mathbf{X} - \left[\mathbf{u}_1^t \ \cdots \ \mathbf{u}_l \ \cdots \mathbf{u}_L^t \right] \cdot \mathbf{Z}^{t+1} \right\|_F^2 \\
= & \left\| \mathbf{X} - \left(\sum_{e \neq l} \mathbf{u}_e^t \left(\mathbf{z}_e^{t+1} \right)^\top + \mathbf{u}_l \left(\mathbf{z}_l^{t+1} \right)^\top \right) \right\|_F^2 \\
= & \left\| \mathbf{R}_l^t - \mathbf{u}_l \left(\mathbf{z}_l^{t+1} \right)^\top \right\|_F^2
\end{aligned}
$$

where $\mathbf{z}_l^\top$ is the l-th row of matrix $\mathbf{Z}$.

Dictionary Update II

- Isolate $\mathbf{R}_l^t$: residual due to atom $\mathbf{u}_l$

$$
\begin{aligned}
& \left\|\mathbf{X} - \left[\mathbf{u}_1^t \ \cdots \ \mathbf{u}_l \ \cdots \mathbf{u}_L^t\right] \cdot \mathbf{Z}^{t+1}\right\|_F^2 \\
= & \left\|\mathbf{X} - \left(\sum_{e \neq l} \mathbf{u}_e^t \left(\mathbf{z}_e^{t+1}\right)^\top + \mathbf{u}_l \left(\mathbf{z}_l^{t+1}\right)^\top\right)\right\|_F^2 \\
= & \left\|\mathbf{R}_l^t - \mathbf{u}_l \left(\mathbf{z}_l^{t+1}\right)^\top\right\|_F^2
\end{aligned}
$$

  where $\mathbf{z}_l^\top$ is the l-th row of matrix $\mathbf{Z}$.
- Changing $\left(\mathbf{z}_l^{t+1}\right)^\top$ while changing $\mathbf{u}_l$ improves the result!

Dictionary Update II

- Isolate $\mathbf{R}_l^t$: residual due to atom $\mathbf{u}_l$

$$
\begin{aligned}
& \left\| \mathbf{X} - \left[\mathbf{u}_1^t \ \cdots \ \mathbf{u}_l \ \cdots \mathbf{u}_L^t \right] \cdot \mathbf{Z}^{t+1} \right\|_F^2 \\
= & \left\| \mathbf{X} - \left(\sum_{e \neq l} \mathbf{u}_e^t \left(\mathbf{z}_e^{t+1} \right)^\top + \mathbf{u}_l \left(\mathbf{z}_l^{t+1} \right)^\top \right) \right\|_F^2 \\
= & \left\| \mathbf{R}_l^t - \mathbf{u}_l \left(\mathbf{z}_l^{t+1} \right)^\top \right\|_F^2
\end{aligned}
$$

  where $\mathbf{z}_l^\top$ is the l-th row of matrix $\mathbf{Z}$.
- Changing $\left(\mathbf{z}_l^{t+1} \right)^\top$ while changing $\mathbf{u}_l$ improves the result!
- How to keep sparseness of representations?

Dictionary Update III

$$\left\| \mathbf{R}_l^t - \mathbf{u}_l \left(\mathbf{z}_l^{t+1}\right)^\top \right\|_F^2$$

where $\mathbf{z}_l^\top$ is the l-th row of matrix $\mathbf{Z}$.

- Changing $\left(\mathbf{z}_l^{t+1}\right)^\top$ while changing $\mathbf{u}_l$ improves the result!
- How to keep sparseness of representations?

Dictionary Update III

$$\left\| \mathbf{R}_l^t - \mathbf{u}_l \left(\mathbf{z}_l^{t+1} \right)^\top \right\|_F^2$$

where $\mathbf{z}_l^\top$ is the l-th row of matrix $\mathbf{Z}$.

- Changing $\left(\mathbf{z}_l^{t+1} \right)^\top$ while changing $\mathbf{u}_l$ improves the result!
- How to keep sparseness of representations?
- Ignore data for which the atom $\mathbf{u}_l$ has coefficient 0, i.e. remove corresponding columns from $\mathbf{R}_l^t$ and entries from $\left(\mathbf{z}_l^{t+1} \right)^\top$.

Dictionary Update III

$$\left\| \mathbf{R}_l^t - \mathbf{u}_l \left(\mathbf{z}_l^{t+1}\right)^\top \right\|_F^2$$

where $\mathbf{z}_l^\top$ is the l-th row of matrix $\mathbf{Z}$.

- Changing $\left(\mathbf{z}_l^{t+1}\right)^\top$ while changing $\mathbf{u}_l$ improves the result!
- How to keep sparseness of representations?
- Ignore data for which the atom $\mathbf{u}_l$ has coefficient 0, i.e. remove corresponding columns from $\mathbf{R}_l^t$ and entries from $\left(\mathbf{z}_l^{t+1}\right)^\top$.
- Obtain new equation:

$$\left\| \mathbf{R}_l'^t - \mathbf{u}_l \left(\mathbf{z}_l'^{t+1}\right)^\top \right\|_F^2$$

Dictionary Update IV

Finding $\mathbf{u}_l^*$:

- $\mathbf{u}_l \left(\mathbf{z}_l'^{t+1}\right)^\top$ is an outer product, i.e. a matrix
- Minimize residual

$$\left\|\mathbf{R}_l'^{t} - \mathbf{u}_l \left(\mathbf{z}_l'^{t+1}\right)^\top\right\|_F^2$$

  by approximating $\mathbf{R}_l'^{t}$ with rank-1 matrix $\mathbf{u}_l \left(\mathbf{z}_l'^{t+1}\right)^\top$
- "Approximately" (Eckart-Young theorem) achieved by SVD of $\mathbf{R}_l'^{t}$:

$$\mathbf{R}_l'^{t} = \tilde{\mathbf{U}}\boldsymbol{\Sigma}\tilde{\mathbf{V}}^\top = \sum_i \sigma_i \tilde{\mathbf{u}}_i \tilde{\mathbf{v}}_i^\top$$

  - $\mathbf{u}_l^* = \tilde{\mathbf{u}}_1$ is first left-singular vector.
  - $\|\mathbf{u}_l^*\|_2 = 1$ naturally satisfied.
- $\left(\mathbf{z}_l'^{t+1}\right)^{\top *} = \sigma_1 \tilde{\mathbf{v}}_1$, recover $\left(\mathbf{z}_l^{t+1}\right)^{\top *}$ by filling in zeros for omitted data

Exercise 1 (blackboard)

Given a signal $\mathbf{x} = (3, 1, -2) \in \mathbb{R}^3$ and an overcomplete dictionary $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4] \in \mathbb{R}^{3\times 4}$,

$$\mathbf{U} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix},$$

find the sparse representation $\mathbf{z}$ of the signal $\mathbf{x}$ with $\|\mathbf{z}\|_0 \leq 2$.

a. Find the atom $\mathbf{u}^{(1)}$ that minimize the reconstruction error $\|\mathbf{x} - \hat{\mathbf{x}}^{(0)}\|_2^2$ where $\hat{\mathbf{x}}^{(0)} = z^{(1)}\mathbf{u}^{(1)}$, and compute the residual $\mathbf{r}^{(1)} = \mathbf{x} - \hat{\mathbf{x}}^{(0)}$.

Exercise 1 (blackboard)

Given a signal $\mathbf{x} = (3, 1, -2) \in \mathbb{R}^3$ and an overcomplete dictionary $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4] \in \mathbb{R}^{3\times 4}$,

$$\mathbf{U} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix},$$

find the sparse representation $\mathbf{z}$ of the signal $\mathbf{x}$ with $\|\mathbf{z}\|_0 \leq 2$.

a. Find the atom $\mathbf{u}^{(1)}$ that minimize the reconstruction error $\|\mathbf{x} - \hat{\mathbf{x}}^{(0)}\|_2^2$ where $\hat{\mathbf{x}}^{(0)} = z^{(1)}\mathbf{u}^{(1)}$, and compute the residual $\mathbf{r}^{(1)} = \mathbf{x} - \hat{\mathbf{x}}^{(0)}$.

b. Find the atom $\mathbf{u}^{(2)}$ that minimize the reconstruction error $\|\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(1)}\|_2^2$ where $\hat{\mathbf{x}}^{(1)} = z^{(2)}\mathbf{u}^{(2)}$.

Exercise 1 (blackboard)

Given a signal $\mathbf{x} = (3, 1, -2) \in \mathbb{R}^3$ and an overcomplete dictionary $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4] \in \mathbb{R}^{3\times 4}$,

$$\mathbf{U} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix},$$

find the sparse representation $\mathbf{z}$ of the signal $\mathbf{x}$ with $\|\mathbf{z}\|_0 \leq 2$.

a. Find the atom $\mathbf{u}^{(1)}$ that minimize the reconstruction error $\|\mathbf{x} - \hat{\mathbf{x}}^{(0)}\|_2^2$ where $\hat{\mathbf{x}}^{(0)} = z^{(1)}\mathbf{u}^{(1)}$, and compute the residual $\mathbf{r}^{(1)} = \mathbf{x} - \hat{\mathbf{x}}^{(0)}$.

b. Find the atom $\mathbf{u}^{(2)}$ that minimize the reconstruction error $\|\mathbf{r}^{(1)} - \hat{\mathbf{x}}^{(1)}\|_2^2$ where $\hat{\mathbf{x}}^{(1)} = z^{(2)}\mathbf{u}^{(2)}$.

c. Write down the sparse representation $\mathbf{z}$ of signal $\mathbf{x}$.

Exercise 4 (blackboard)

a. Find an overcomplete dictionary and a vector $\mathbf{x}$ such that the approximation $\hat{\mathbf{x}}$ resulting from the matching pursuit algorithm will never exactly equal $\mathbf{x}$ no matter the number of iterations. Note that this implies that at least one atom will be selected multiple times.

b. Find an instance where the sparse representation returned by matching pursuit (assuming that after some number of iterations, the approximation is perfect) is not optimal, i.e. there is a different representation for which the 0-norm is strictly smaller.