

Solution for Assignment 8

1. a) For the first datapoint we get $t_1 = 1$ and $b_1 = 2$. Hence, we get the equation $\alpha_1 + \alpha_0 = 2$ from $f(t_1) = b_1$. We proceed analogously for all $i \in [5]$ and obtain the system of linear equations

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 6 \\ 8 \end{bmatrix}.$$

- b) Since there are 5 equations and only 2 unknowns, we generally don't expect this system to have a solution. In particular, if our data comes from a real world process we cannot expect the data to exactly lie on a line (even if this is a good model for the real world process).

- c) Let A be the system matrix from the linear system above and let \mathbf{b} be the right-hand side. We then get

$$A^\top A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix}$$

and

$$A^\top \mathbf{b} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 87 \\ 24 \end{bmatrix}.$$

Hence, we get the normal equations

$$\begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} 87 \\ 24 \end{bmatrix}.$$

We can now solve this for α_1 and α_0 using Gauss-elimination or other techniques. One way to do it in this particular case, is to subtract the second row 3 times from the first row to get

$$\begin{bmatrix} 10 & 0 \\ 15 & 5 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} 15 \\ 24 \end{bmatrix}.$$

We then conclude $\alpha_1 = \frac{3}{2}$ and $\alpha_0 = \frac{3}{10}$. In particular, we conclude that the line $f(t) = \frac{3}{2}t + \frac{3}{10}$ explains our data best (among all lines).

2. a) We want to find α_0, α_1 such that

$$\sum_{k=1}^m \lambda_k (b_k - (\alpha_0 + \alpha_1 t_k))^2$$

is minimized. We rewrite this to

$$\sum_{k=1}^m \lambda_k (b_k - (\alpha_0 + \alpha_1 t_k))^2 = \sum_{k=1}^m (\sqrt{\lambda_k} b_k - \sqrt{\lambda_k} (\alpha_0 + \alpha_1 t_k))^2 = \left\| \Lambda^{\frac{1}{2}} \mathbf{b} - \Lambda^{\frac{1}{2}} A \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \right\|^2,$$

where

$$A^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_m} \end{bmatrix}.$$

Using Fact 5.3.1 with $\mathbf{b}' := A^{\frac{1}{2}}\mathbf{b}$ and $A' := A^{\frac{1}{2}}A$, we therefore get that the optimal choice for α_0, α_1 satisfies

$$(A')^\top A' \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = A'^\top \mathbf{b}'.$$

Now observe that $A'^\top A' = A^\top (A^{\frac{1}{2}})^\top A^{\frac{1}{2}} A = A^\top \Lambda A$ which is invertible by assumption. Thus, we conclude that

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = (A^\top \Lambda A)^{-1} A^\top \Lambda \mathbf{b}.$$

- b)** By Lemma 5.2.4, $A^\top \Lambda A$ is invertible if and only if $A^{\frac{1}{2}}A$ has linearly independent columns. Similarly, $A^\top A$ is invertible if and only if A has linearly independent columns. Thus, it suffices to prove that $A^{\frac{1}{2}}A$ has linearly independent columns if and only if A has linearly independent columns. Observe that $A^{\frac{1}{2}}$ has inverse

$$A^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sqrt{\lambda_m}} \end{bmatrix}$$

since $\lambda_i > 0$ for all $i \in [m]$. We can use this to get

$$A\mathbf{x} = \mathbf{0} \iff A\mathbf{x} = A^{-\frac{1}{2}}\mathbf{0} \iff A^{\frac{1}{2}}A\mathbf{x} = \mathbf{0}$$

which directly implies that A has linearly independent columns if and only if $A^{\frac{1}{2}}A$ has linearly independent columns.

- 3.** The linear transformation given by $A = I - 2\mathbf{v}\mathbf{v}^\top$ corresponds to reflection along the hyperplane $H = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \cdot \mathbf{v} = 0\}$. To see this, recall from Theorem 5.2.6 that the projection matrix for projection onto $\text{Span}(\mathbf{v})$ is given by $B = \frac{\mathbf{v}\mathbf{v}^\top}{\|\mathbf{v}\|^2}$. Since \mathbf{v} is a unit vector, this simplifies to $B = \mathbf{v}\mathbf{v}^\top$. In particular, this means that $A = I - 2B$ applied to some vector $\mathbf{x} \in \mathbb{R}^2$ will subtract the projection of \mathbf{x} onto $\text{Span}(\mathbf{v})$ twice from \mathbf{x} itself. In other words, assume we split \mathbf{x} into two parts $\mathbf{x} = \mathbf{x}_\parallel + \mathbf{x}_\perp$ with $\mathbf{x}_\parallel = B\mathbf{x}$ and $\mathbf{x}_\perp \cdot \mathbf{v} = 0$. Then the transformation given by matrix A maps \mathbf{x} to the vector

$$A\mathbf{x} = (I - 2B)(\mathbf{x}_\parallel + \mathbf{x}_\perp) = \mathbf{x}_\parallel + \mathbf{x}_\perp - 2B\mathbf{x}_\parallel - 2B\mathbf{x}_\perp = \mathbf{x}_\parallel + \mathbf{x}_\perp - 2\mathbf{x}_\parallel - 2\mathbf{v}\mathbf{v}^\top \mathbf{x}_\perp = \mathbf{x}_\perp - \mathbf{x}_\parallel - \mathbf{0} = \mathbf{x}_\perp - \mathbf{x}_\parallel.$$

A picture of this is provided in Figure 1.

- 4. a)** We obtained the new datapoints from the old one by shifting them along the t -axis. Imagine a line that optimally fits the old datapoints. By shifting this line the same amount in the direction of the t -axis, we should get a line that optimally fits the new datapoints. In other words, it should be possible to obtain the optimal line for the new datapoints from the optimal line for the old datapoints by a shift along the t -axis. This should also work the other way around, i.e. given an optimal line for the new datapoints we should be able to shift the line by $-c$ in the direction of the t -axis to get an optimal line for the old datapoints.

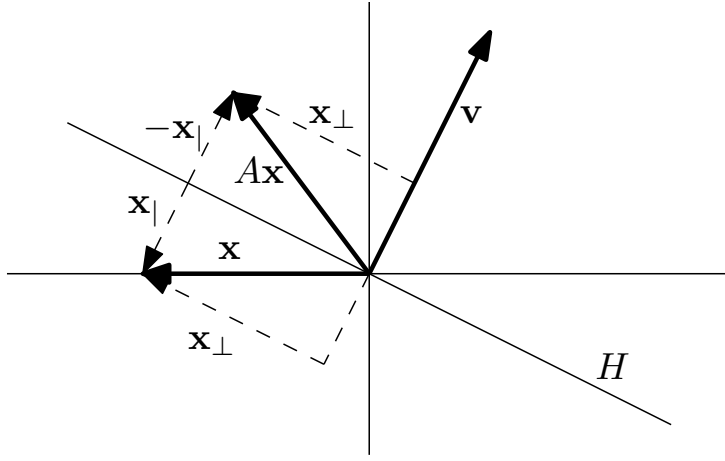


Figure 1: A sketch of the transformation.

It remains to think about this shifting operation in terms of the parameters α and α' . The main insight here is that shifting a line does not change its slope. Hence we expect to have $\alpha_1 = \alpha'_1$ but not $\alpha_0 = \alpha'_0$.

b) We compute the scalar product and check that it yields zero. Concretely, we have

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \begin{bmatrix} t'_1 \\ \vdots \\ t'_m \end{bmatrix} = \sum_{k=1}^m t'_m = \sum_{k=1}^m (t_m + c) = \sum_{k=1}^m (t_k - \frac{1}{m} \sum_{i=1}^m t_i) = \sum_{k=1}^m t_k - m \frac{1}{m} \sum_{i=1}^m t_i = \sum_{k=1}^m t_k - \sum_{i=1}^m t_i = 0$$

and hence the columns of A' are indeed orthogonal.

c) We prove that for all $\alpha' \in \mathbb{R}^2$ and $\alpha = \alpha' + \begin{bmatrix} c\alpha'_1 \\ 0 \end{bmatrix}$, we have $\|A'\alpha' - \mathbf{b}\|^2 = \|A\alpha - \mathbf{b}\|^2$. From this, it follows that α' is optimal if and only if α is optimal.

Thus, consider an arbitrary $\alpha' \in \mathbb{R}^2$ and let $\alpha = \alpha' + \begin{bmatrix} c\alpha'_1 \\ 0 \end{bmatrix}$. We compute

$$\begin{aligned} \|A'\alpha' - \mathbf{b}\|^2 &= \sum_{k=1}^m (b_k - (\alpha'_0 + \alpha'_1 t'_k))^2 \\ &= \sum_{k=1}^m (b_k - (\alpha'_0 + \alpha'_1 (t_k + c)))^2 && (t'_k = t_k + c) \\ &= \sum_{k=1}^m (b_k - (\alpha'_0 + \alpha'_1 c + \alpha'_1 t_k))^2 \\ &= \sum_{k=1}^m (b_k - (\alpha_0 + \alpha'_1 t_k))^2 && (\alpha_0 = \alpha'_0 + c\alpha'_1) \\ &= \sum_{k=1}^m (b_k - (\alpha_0 + \alpha_1 t_k))^2 && (\alpha_1 = \alpha'_1) \\ &= \|A\alpha - \mathbf{b}\|^2 \end{aligned}$$

and therefore conclude the proof.

d) By the closed form solution and subtask b), we get that the optimal α' satisfies

$$\begin{bmatrix} \alpha'_0 \\ \alpha'_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{m} \sum_{k=1}^m b_k \\ (\sum_{k=1}^m t'_k b_k) / (\sum_{k=1}^m t_k^2) \end{bmatrix}.$$

By subtask c), the optimal α is hence given by

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{m} \sum_{k=1}^m b_k + c(\sum_{k=1}^m t'_k b_k) / (\sum_{k=1}^m t_k'^2) \\ (\sum_{k=1}^m t'_k b_k) / (\sum_{k=1}^m t_k'^2) \end{bmatrix}$$

where $c = -\frac{1}{m} \sum_{k=1}^m t_k$ and $t'_k = t_k + c$. This is a closed form solution because the right-hand side can be directly computed from the datapoints.

5. Consider an arbitrary linear combination

$$\sum_{i=1}^n \alpha_i \mathbf{u}_i + \sum_{i=1}^m \beta_i \mathbf{w}_i = \mathbf{0}$$

with scalars $\alpha_1, \dots, \alpha_n$ and β_1, \dots, β_m . Observe that we must have

$$\sum_{i=1}^n \alpha_i \mathbf{u}_i \in U \text{ and } -\sum_{i=1}^m \beta_i \mathbf{w}_i \in W$$

since U and W are vector spaces (Lemma 4.12) and linear combinations taken in a vector space are again contained in that vector space (Lemma 4.14). But we also know that

$$\sum_{i=1}^n \alpha_i \mathbf{u}_i = -\sum_{i=1}^m \beta_i \mathbf{w}_i$$

which, by our assumption $U \cap W = \{\mathbf{0}\}$, implies

$$\sum_{i=1}^n \alpha_i \mathbf{u}_i = \sum_{i=1}^m \beta_i \mathbf{w}_i = \mathbf{0}.$$

Since $\mathbf{u}_1, \dots, \mathbf{u}_n$ are linearly independent, we conclude $\alpha_1 = \dots = \alpha_n = 0$. Similarly, we get $\beta_1 = \dots = \beta_m = 0$. We conclude that the only way of expressing $\mathbf{0}$ as a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_n, \mathbf{w}_1, \dots, \mathbf{w}_m$ is the trivial linear combination, and thus these $m+n$ vectors are linearly independent.

6. Using Theorem 5.1.7 and Corollary 5.1.9, we can decompose $\mathbf{x} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in \mathbf{C}(A^\top)$ and $\mathbf{w} \in \mathbf{N}(A)$. Now observe that

$$A\mathbf{v} = A\mathbf{v} + A\mathbf{w} = A\mathbf{x} = \mathbf{b}$$

since $\mathbf{w} \in \mathbf{N}(A)$. With $\mathbf{x}^* = \mathbf{v}$ this proves existence and it remains to argue uniqueness. Indeed, any distinct $\mathbf{x}, \mathbf{y} \in \mathbf{C}(A^\top)$ with $A\mathbf{x} = \mathbf{b} = A\mathbf{y}$ must satisfy $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. This means that $\mathbf{x} - \mathbf{y} \in \mathbf{N}(A)$. But since $\mathbf{x}, \mathbf{y} \in \mathbf{C}(A^\top)$, we also get $\mathbf{x} - \mathbf{y} \in \mathbf{C}(A^\top)$ and thus $\mathbf{x} = \mathbf{y}$. We conclude that \mathbf{x}^* is unique.