Exercise A: Least square problems

A1

Proof. Minimising $\|Ax - b\|_2$ with respect to x means finding x such that the derivative of $\|Ax - b\|_2$ with respect to x is equal to zero. Equivalently, we can minimise the square of the norm which can be developed as follows:

$$\|Ax - b\|_2^2 = \|Ax\|_2^2 - 2\langle Ax, b \rangle + \|b\|_2^2$$
$$= x^\top A^\top A x - 2 x^\top A^\top b + b^\top b.$$

The derivative of this expression with respect to x is

$$\frac{\partial \|Ax - b\|_2^2}{\partial x} = \frac{\partial (x^\top A^\top A x - 2 x^\top A^\top b + b^\top b)}{\partial x}$$
$$= \frac{\partial (x^\top A^\top A x)}{\partial x} - \frac{\partial (2 x^\top A^\top b)}{\partial x} + \frac{\partial (b^\top b)}{\partial x}$$
$$= 2 A^\top A x - 2 A^\top b + 0.$$

We set this derivative to zero to obtain the following:

$$\frac{\partial \|Ax - b\|_2^2}{\partial x} = 0$$
$$\iff 2 A^\top A x - 2 A^\top b = 0$$
$$\iff 2 A^\top A x = 2 A^\top b$$
$$\iff A^\top A x = A^\top b.$$

It is indeed a minimum and not a maximum, as we consider $\|Ax - b\|_2^2$, which is a quadratic expression having only one extremum that is a minimum.

To prove that when A has full column rank, i.e. $\mathrm{rank}(A) = n$, the solution is unique, we first show that $\mathrm{Ker}(A^\top A) = \mathrm{Ker}(A)$:

$$\forall x \in \mathrm{Ker}(A) : Ax = 0 \iff A^\top A x = 0 \iff x \in \mathrm{Ker}(A^\top A) \implies \mathrm{Ker}(A) \subseteq \mathrm{Ker}(A^\top A).$$

$$\forall x \in \mathrm{Ker}(A^\top A) : A^\top A x = 0 \iff x^\top A^\top A x = 0 \iff \|Ax\|_2 = 0 \iff Ax = 0 \iff x \in \mathrm{Ker}(A)$$
$$\implies \mathrm{Ker}(A^\top A) \subseteq \mathrm{Ker}(A).$$

Combining these results, we have $\mathrm{Ker}(A^\top A) = \mathrm{Ker}(A)$.

According to the rank-nullity theorem,

$$\mathrm{rank}(A) = n - \dim(\mathrm{Ker}(A)).$$

However, $\mathrm{rank}(A) = n$, and thus

$$n = n - \dim(\mathrm{Ker}(A))$$
$$\iff \dim(\mathrm{Ker}(A)) = 0 = \dim(\mathrm{Ker}(A^\top A))$$
$$\implies \mathrm{Ker}(A^\top A) = \{0\}.$$

As $\mathrm{Ker}(A^\top A) = \{0\}$ and $A^\top A$ is square, we deduce that $A^\top A$ is invertible. It then follows from Theorem 2.1 of the lecture notes that the solution of the system is unique. $\qquad\square$

A2

Suppose the QR decomposition of A is given by $Q \begin{pmatrix} R \\ 0 \end{pmatrix}$, where $Q \in \mathbb{R}^{m \times m}$ is unitary and $R \in \mathbb{R}^{n \times n}$ is upper triangular. We will express the solution of

$$A^\top A x = A^\top b \tag{1}$$

in terms of the QR decomposition of A.

Let $R_f = \begin{pmatrix} R \\ 0 \end{pmatrix}$. We can rewrite (1) as

$$(QR_f)^\top QR_f x = (QR_f)^\top b$$
$$R_f^\top Q^\top QR_f x = R_f^\top Q^\top b.$$

As Q is unitary and hence $Q^\top Q = I$, we have

$$\begin{pmatrix} R^\top & 0 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} x = \begin{pmatrix} R^\top & 0 \end{pmatrix} Q^\top b.$$

If we call $\hat{Q}$ the matrix consisting of the first n columns of Q, (1) becomes

$$R^\top R x = R^\top \hat{Q}^\top b.$$

From Theorem 2.8 of the lecture notes, we know that every matrix $A \in \mathbb{C}^{m \times n}$ of full column rank admits a factorization $A = Q_1 R_1$ where $Q_1 \in \mathbb{C}^{m \times n}$ is an isometry and $R_1 \in \mathbb{C}^{n \times n}$ is an upper triangular matrix with positive diagonal. The matrix Q_1 corresponds to $\hat{Q}$ and R_1 simply to R. Hence, we deduce that R is invertible. This allows us to premultiply both sides of the equation by the inverse of $R^\top$:

$$R^{-\top} R^\top R x = R^{-\top} R^\top \hat{Q}^\top b$$
$$R x = \hat{Q}^\top b.$$

The solution of (1) is therefore $x = R^{-1} \hat{Q}^\top b$ and the computation of the solution is reduced to the resolution of a single triangular system of linear equations (which can be solved efficiently using backward substitution).

Exercise B: Low-rank approximation

B1

For every matrix $A \in \mathbb{R}^{m \times n}$, there exist unitary transformations $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that

$$A = U\Sigma V^*, \quad \text{where} \quad \Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0_{r \times (n-r)} \\ 0 & & \sigma_r & \\ \hline & 0_{(m-r) \times r} & & 0_{(m-r) \times (n-r)} \end{array} \right],$$

with real positive singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$. These singular values are unique. Indeed, they can be derived from the eigenvalues of the Hermitian matrix $A^* A$ (as explained on p. 39 of the course notes) which are uniquely defined.

Next, we show that the rank of a matrix is equal to its number of nonzero singular values.

Proof. We first prove that if A is a $k \times \ell$ matrix and B is a full-rank $\ell \times \ell$ matrix then $\text{rank}(AB) = \text{rank}(A)$. Let S_1 be the space generated by the columns of A and S_2 the space generated by the columns of AB. If $x \in S_1$ then there exists $y \in \mathbb{R}^{\ell \times 1}$ such that $x = Ay$. Considering $\hat{y} = B^{-1}y$, we note that $x = ABB^{-1}y = (AB)\hat{y}$ and so $x \in S_2$. If $x \in S_2$ then there exists $y \in \mathbb{R}^{\ell \times 1}$ such that $x = (AB)y$. We note that $x = (A)By$ and so $x \in S_2$. This shows that the spaces generated by the columns of A and AB coincide and hence $\text{rank}(A) = \text{rank}(AB)$.

Similarly we prove that if A is a $k \times \ell$ matrix and B is a full-rank $k \times k$ matrix then $\text{rank}(BA) = \text{rank}(A)$. Let S_1 be the space generated by the rows of A and S_2 the space generated by the rows of BA. If $x \in S_1$ then there exists $y \in \mathbb{R}^{1 \times k}$ such that $x = yA$. Considering $\hat{y} = yB^{-1}$ we note that $x = yB^{-1}BA = \hat{y}(BA)$ and so $x \in S_2$. If $x \in S_2$ then there exists $y \in \mathbb{R}^{1 \times k}$ such that $x = y(BA)$. We note that $x = yB(A)$ and so $x \in S_1$. This shows that the spaces generated by the rows of A and AB coincide and hence $\text{rank}(A) = \text{rank}(BA)$.

We know that the rank of a diagonal matrix is equal to the number of its nonzero entries. We also note that in the decomposition $A = U\Sigma V^*$, U and V^* have full rank. Therefore, $\text{rank}(A) = \text{rank}(\Sigma) = r$ where r is the number of nonzero singular values. $\square$

B2

Let $x \in \mathbb{R}^{m \times n}$ be such that $|X_{ij}| \leqslant \varepsilon$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. Let $\|X\|_2$ be the 2-norm of X and let $\|X\|_F$ be its Frobenius norm. We show that $\|X\|_2 \leqslant \|X\|_F \leqslant \sqrt{mn}\varepsilon$.

Proof. First, we show the first inequality. We know from the lecture notes that

$$\|X\|_2 = \sigma_{\max},$$

$$\|X\|_F = \left[\sum_i \sigma_i^2\right]^{1/2},$$

where σ_i are the singular values of X. From this, it is immediately clear that $\|X\|_2 \leqslant \|X\|_F$.

Next, we use an equivalent form of the Frobenius norm to show the second inequality:

$$\|X\|_F = \left[\sum_{i,j} |X_{ij}|^2\right]^{1/2}.$$

Knowing that $|X_{ij}| \leqslant \varepsilon$, it is immediate that $\|X\|_F \leqslant \left[\sum_{i,j} \varepsilon^2\right]^{1/2} = \left[mn\varepsilon^2\right]^{1/2} = \sqrt{mn}\varepsilon$. This concludes the proof. $\square$

We also give an example where these bounds are tight. Indeed, consider the matrix $X = \varepsilon \in \mathbb{R}^{1 \times 1}$. Clearly, we have $|X_{ij}| \leqslant \varepsilon$ for all i, j (only one value is possible for each). We know that the only singular value of this matrix is ε, and hence

$$\|X\|_2 = \|X\|_F = \sqrt{1 \cdot 1}\varepsilon = \varepsilon.$$

B3

To answer this question, we will need to use the following lemma L1:

$$\sigma_{k+l+1}(A + X) \leqslant \sigma_{k+1}(A) + \sigma_{l+1}(X). \tag{2}$$

Proof. To prove this inequality, we will use Theorem 3.28 in the lecture notes :

$$\sigma_{n-j+1}(B) = \min_{\mathcal{S}_j} \max_{\boldsymbol{x} \in \mathcal{S}_j \setminus \{0\}} \frac{\|B\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2},$$

with $n = \min(m, n)$ without loss of generality.

Now, we are interested in the singular values of a sum of two matrices and we set $k + l = n - j \iff j = n - k - l$.

$$\sigma_{k+l+1}(A + X) = \min_{\mathcal{S}_{n-k-l}} \max_{\boldsymbol{x} \in \mathcal{S}_{n-k-l} \setminus \{0\}} \frac{\|(A + X)\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}.$$

We also have

$$\sigma_{k+1}(A) = \min_{\mathcal{S}_{n-k}} \max_{\boldsymbol{x} \in \mathcal{S}_{n-k} \setminus \{0\}} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$$

$$\sigma_{l+1}(X) = \min_{\mathcal{S}_{n-l}} \max_{\boldsymbol{x} \in \mathcal{S}_{n-l} \setminus \{0\}} \frac{\|X\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}.$$

Let us denote by V_A the subset of dimension $n - k$ in which we reach the minimum of the first equality, and V_X the subset of dimension $n - l$ in which we reach the minimum of the second equality.

If we take the intersection $W = V_A \cap V_X$, the dimension of W is at least $n - l - k$ (by applying the identity $\dim(U \cap V) = \dim(U) + \dim(V) - \dim(U \cup V)$, observing that $\dim(U \cup V) \leqslant n$) and by definition we have $W \subseteq V_A$, $W \subseteq V_X$.

Moreover, when $p \leqslant q$, we have $\max_{\boldsymbol{x} \in \mathcal{S}_p} \|M\boldsymbol{x}\|_2 \leqslant \max_{\boldsymbol{x} \in \mathcal{S}_q} \|M\boldsymbol{x}\|_2 \quad \forall \mathcal{S}_p \subseteq \mathcal{S}_q.$

That means, in our case,

$$\max_{\boldsymbol{x}\in W} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leqslant \max_{\boldsymbol{x}\in V_A} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}$$

$$\max_{\boldsymbol{x}\in W} \frac{\|X\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leqslant \max_{\boldsymbol{x}\in V_X} \frac{\|X\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2 n}.$$

With this, we have all the tools to write the following:

$$
\begin{aligned}
\sigma_{k+l+1}(A + X) &= \min_{\mathcal{S}_{n-k-l}} \max_{\boldsymbol{x}\in\mathcal{S}_{n-k-l}\backslash\{0\}} \frac{\|(A + X)\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \\
&\leqslant \min_{\mathcal{S}_{n-k-l}} \left(\max_{\boldsymbol{x}\in\mathcal{S}_{n-k-l}\backslash\{0\}} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} + \frac{\|X\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \right) \\
&\leqslant \max_{\boldsymbol{x}\in W\backslash\{0\}} \left(\frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} + \frac{\|X\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \right) \\
&\leqslant \max_{\boldsymbol{x}\in V_A\backslash\{0\}} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} + \max_{\boldsymbol{x}\in V_X\backslash\{0\}} \frac{\|X\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \\
&\leqslant \sigma_{k+1}(A) + \sigma_{l+1}(X).
\end{aligned}
$$

From the second inequality to the third, we chose $\mathcal{S}_{n-k-l}$ to be a subset of W of dimension $(n - k - l)$ and then, of course, the whole expression with the minimum is less than or equal to the case when $\mathcal{S}_{n-k-l}$ is that subset particularly, as it is a minimum. □

With this inequality, we want to know what happens to the smallest $(\min(m,n) - r)$ singular values of B with r, the rank of A. We take $k = r$ and $l = 0$, and we have

$$\sigma_{r+1}(B) \leqslant \sigma_{r+1}(A) + \sigma_1(X) \tag{3}$$
$$\leqslant 0 + \sqrt{mn}\epsilon. \tag{4}$$

Varying l from 0 to $(\min(m,n) - r)$, we obtain that

$$\sigma_{r+i}(B) \leqslant \sqrt{mn}\epsilon \qquad i = 1, \ldots, (\min(m,n) - r).$$

Indeed, we used the upper bound found in B2 such that $\sigma_i(X) \leqslant \sigma_1(X) = \|X\|_2 \leqslant \sqrt{mn}\epsilon$.
Next, we want to know the rank of A based on the matrix B. Once again, we will use the lemma L1.

$$\sigma_{k+l+1}(A) = \sigma_{k+l+1}(B + (-X)) \leqslant \sigma_{k+1}(B) + \sigma_{l+1}(-X)$$

We know that the elements of $(-X)$ follow $|-X_{ij}| = |X_{ij}| \leqslant \epsilon$ and so we can use the upper bound in B2. Taking $k = r - 1$ and $l = 0$, we have

$$
\begin{aligned}
\sigma_r(A) &\leqslant \sigma_r(B) + \sigma_1(-X) \\
&\leqslant \sigma_r(B) + \sqrt{mn}\epsilon \\
\Longleftrightarrow \sigma_r(B) &\geqslant \sigma_r(A) - \sqrt{mn}\epsilon \\
\sigma_r(B) &> \sqrt{mn}\epsilon.
\end{aligned}
$$

For the last inequality, we used the lower bound $\sigma_r(A) > 2\sqrt{mn}\epsilon$.

A criterion that can then be used to estimate the rank r of A is to take the smallest r such that $\sigma_{r+1}(B) \leqslant \sqrt{mn}\varepsilon$. Indeed, this will work as we have $\sigma_r(B) > \sqrt{mn}\epsilon$. This is very similar to the description given on p. 60 of the lecture notes, concerning the numerical rank of the matrix A.

Exercise C: Realization theory

In this last exercise, we are interested in finding an AR model corresponding to the data obtained during the coronavirus pandemic. Indeed, we want to find the parameters α_i of

$$y(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i y(t - i)$$

Let us rewrite the problem as a system of linear equations:

$$
\begin{pmatrix}
1 & y(p-1) & y(p-2) & \dots & y(0) \\
1 & y(p) & y(p-1) & \dots & y(1) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & y(N-1) & y(N-2) & \dots & y(N-p)
\end{pmatrix}
\begin{pmatrix}
\alpha_0 \\
\alpha_1 \\
\vdots \\
\alpha_p
\end{pmatrix}
=
\begin{pmatrix}
y(p) \\
\vdots \\
y(N)
\end{pmatrix}
\tag{5}
$$

Note that we consider $\hat{y}(0) = y(0), \dots, \hat{y}(p-1) = y(p-1)$ as initial conditions.

We want to solve the system (5) to obtain an estimate of the optimal parameters α_i. We call A the matrix and we observe that it usually has more lines than columns. Therefore, it is of interest to solve the system using the least squares method. To do so, we can use what we obtained in question A2 to solve the problem with a QR decomposition and backward substitution. Indeed, we must simply solve the triangular system :

$$
Rx = \hat{Q}^\top b,
$$

where $\hat{Q}$ and R are derived from the QR decomposition of A.

We consider two cases: the lockdown mode and the social distancing mode. Moreover, for each case, we divide the data set into a training set (70 % of the data) and a validation set (30 % of the data).

Considering one mode, (5) has to be solved using the data of the training set. Next, an estimate y that we call $\hat{y}$ can be reconstructed using the AR model equation. The first p components of $\hat{y}$ are given by the initial conditions and then the next components are constructed based on the previously computed estimates:

$$
\hat{y}(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \hat{y}(t-i).
$$

The error between the true y and the estimate $\hat{y}$ can be computed on both the training and validation set. This allows one to judge the quality of the model.

Next, we show the results we obtained while varying the parameter p.

Lockdown mode   First, we study the lockdown mode. As one can see in Figure 1, when the parameter p increases, the training error has a tendency to decrease. Indeed, the AR model that we obtain is able to better represent our training set data as the size of the history grows.

Then, in Figure 2, we observe that the validation error decreases (approximatively) until $p = 19$, and from then on, the error rises again. Through this, we can understand that around $p = 20$, the model starts "overfitting" our training set, and so it gives a good approximation for the training set but performs poorly on the validation set.

Social distancing mode   In the social distancing mode, we can make the same observations (in Figures 3 and 4) as in the previous mode: the error in the training set decreases when p increases except at some peaks. For the validation set, we once again observe overfitting when p is higher than 19.

Discussion

In both modes, we have confirmed our intuition that the error in the training set should decrease. However, it is not monotonic and we still observe some peaks. One way to try to explain this is that when we solve the least squares problem, we minimize the error between the real values $y(t)$ and $\tilde{y}(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i y(t-i)$, that is, $e_1 = \|y - \tilde{y}\|_2$. However, when we compute the error on the training set, we consider $\hat{y}(t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \hat{y}(t-i)$, that is, $e_2 = \|y - \hat{y}\|_2$, and so we propagate the errors along with the time t. Clearly, this last error e_2 is bigger than the first one, e_1. Our intuition that the error should decrease monotonically can only apply to the error e_1 (which we have indeed observed in Figure 5) and not to e_2.

Then, about the validation set errors, we can conclude that overfitting a training set does not yield good results for future predictions.

Note that in the analysis above, we only considered values of p such that the linear system of equations has more rows than columns. In the case that we have more columns than rows, our system is underdetermined and so solving the system yields an infinity of solutions that are exact (if there are no linear combinations in the rows of course) which means that our training error is zero in theory (in practice, we have errors of order 10^{-4}, for the confinement mode, and 10^{-9} for the social distancing mode, due to numerical imprecisions). As expected, we still have an overfitting that is visible in the validation set errors as they are very big for both modes.
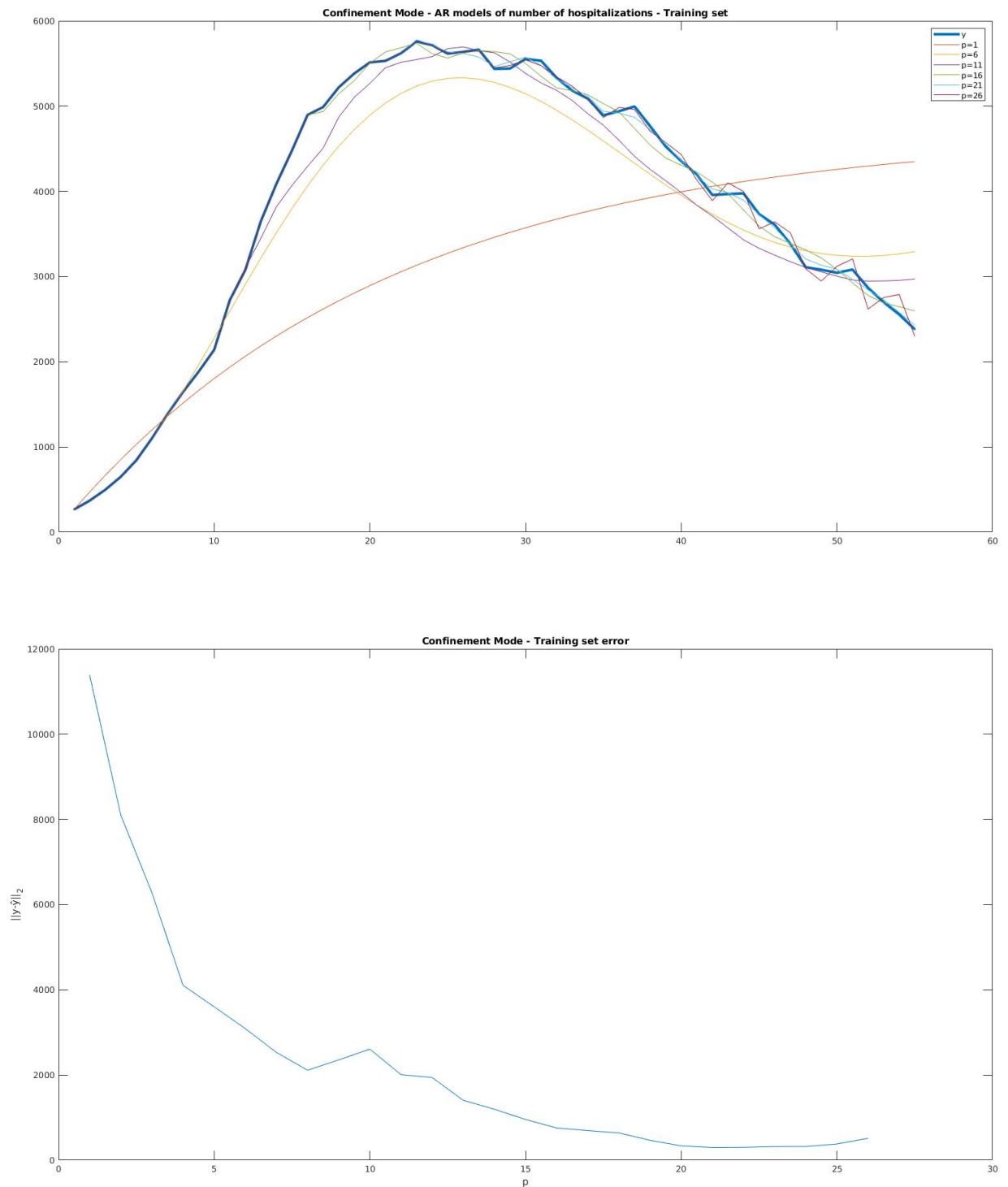
Figure 1: Lockdown mode: training set predictions and error using the AR model, for different values of p.

Bonus question

If we approximate the model in real time with data that we receive at fixed intervals, with our approach we will have to solve a new "full-size" least squares problem (as defined at the beginning of this section) everytime we want to include new data to improve the system. Indeed, we will have to add new rows each time we receive new data and solve it again. This can be expensive especially if the system is big.

That is why maintaining a real-time model with our approach to solve least squares problems is sub-optimal. It would be better if we could only consider the small changes without having to solve the whole system, for example by considering a recursive method to solve least squares problems.
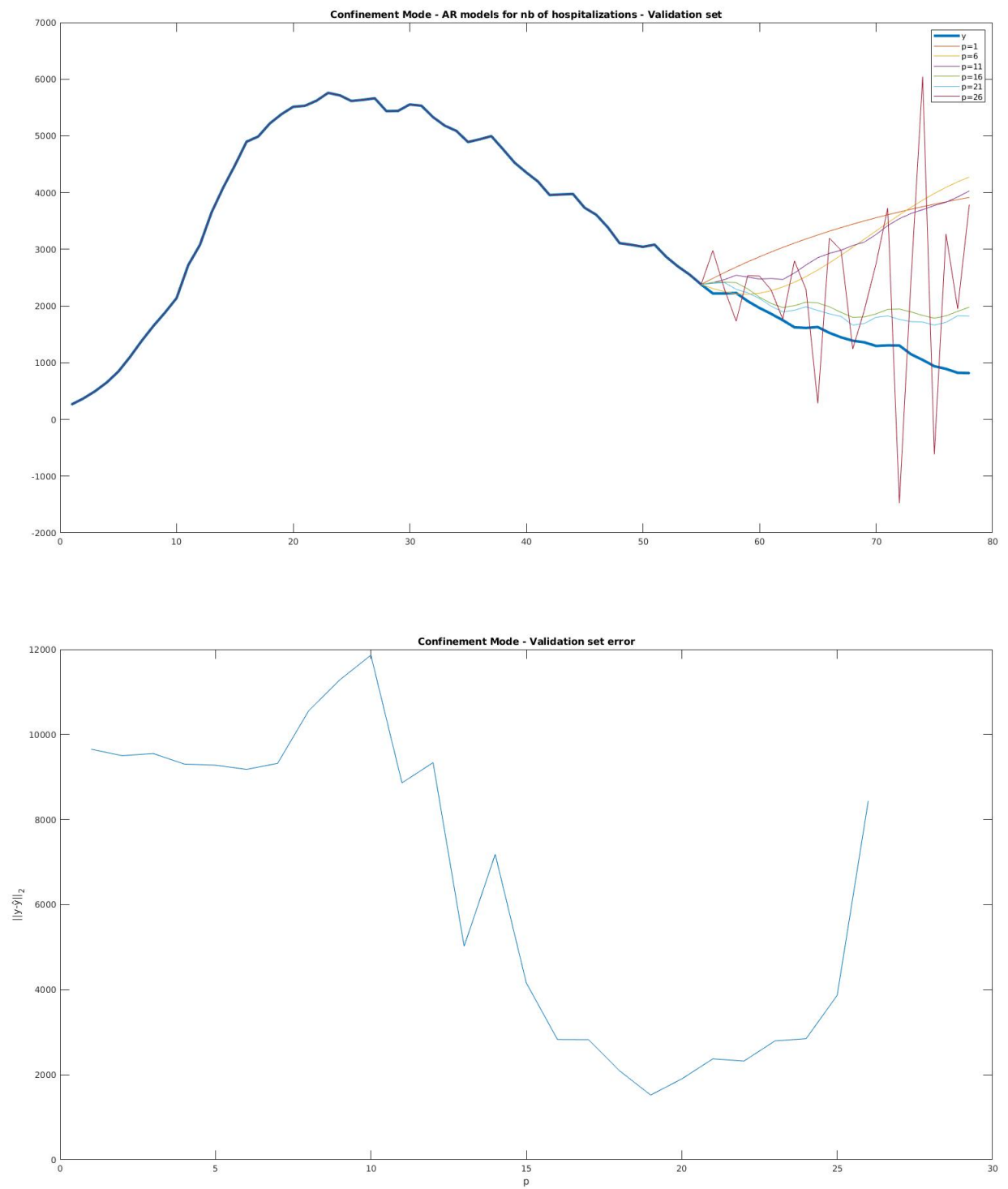
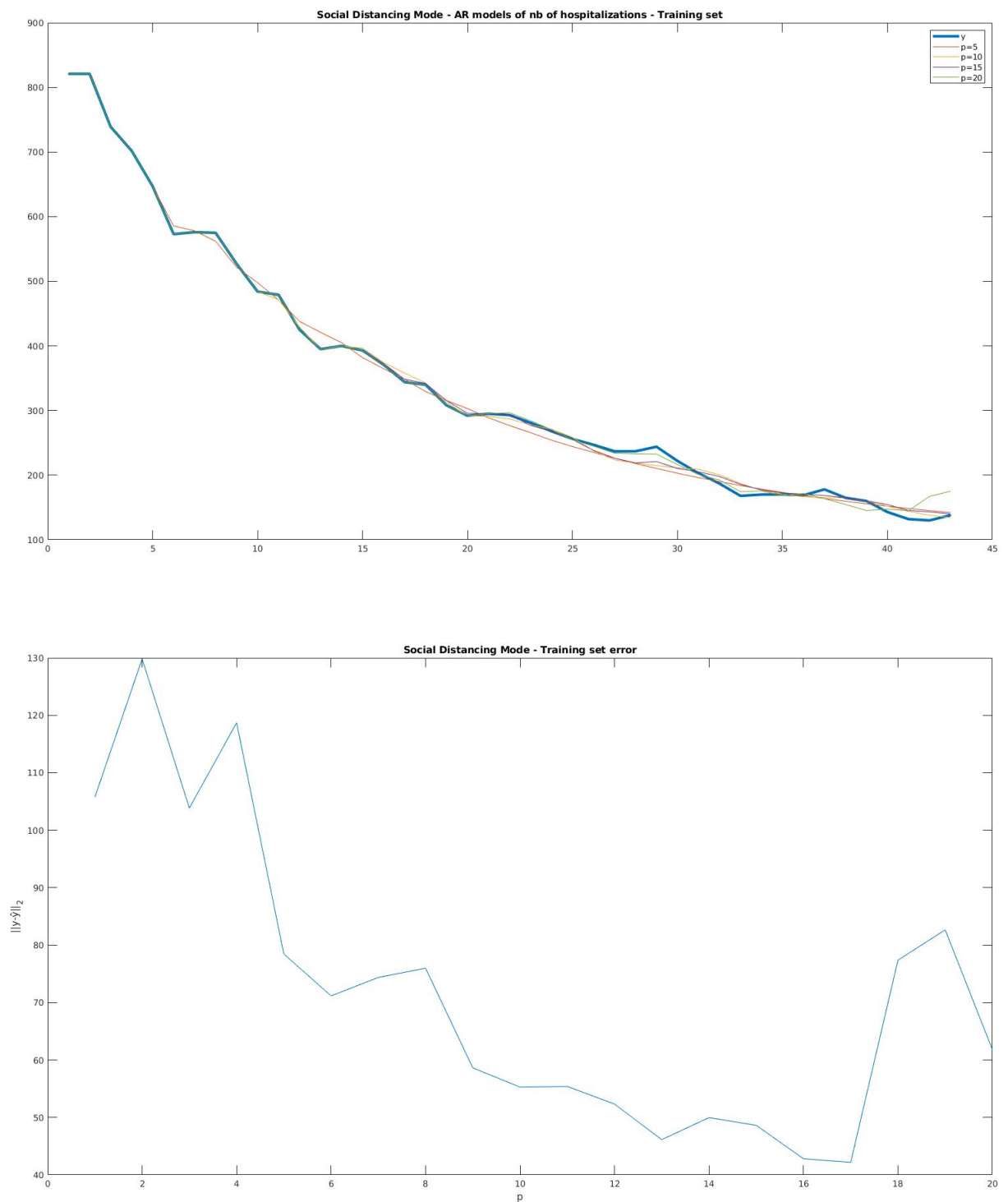Figure 2: Lockdown mode: validation set predictions and error using the AR model, for different values of p.

Figure 3: Social distancing mode: training set predictions and error using the AR model, for different values of p.
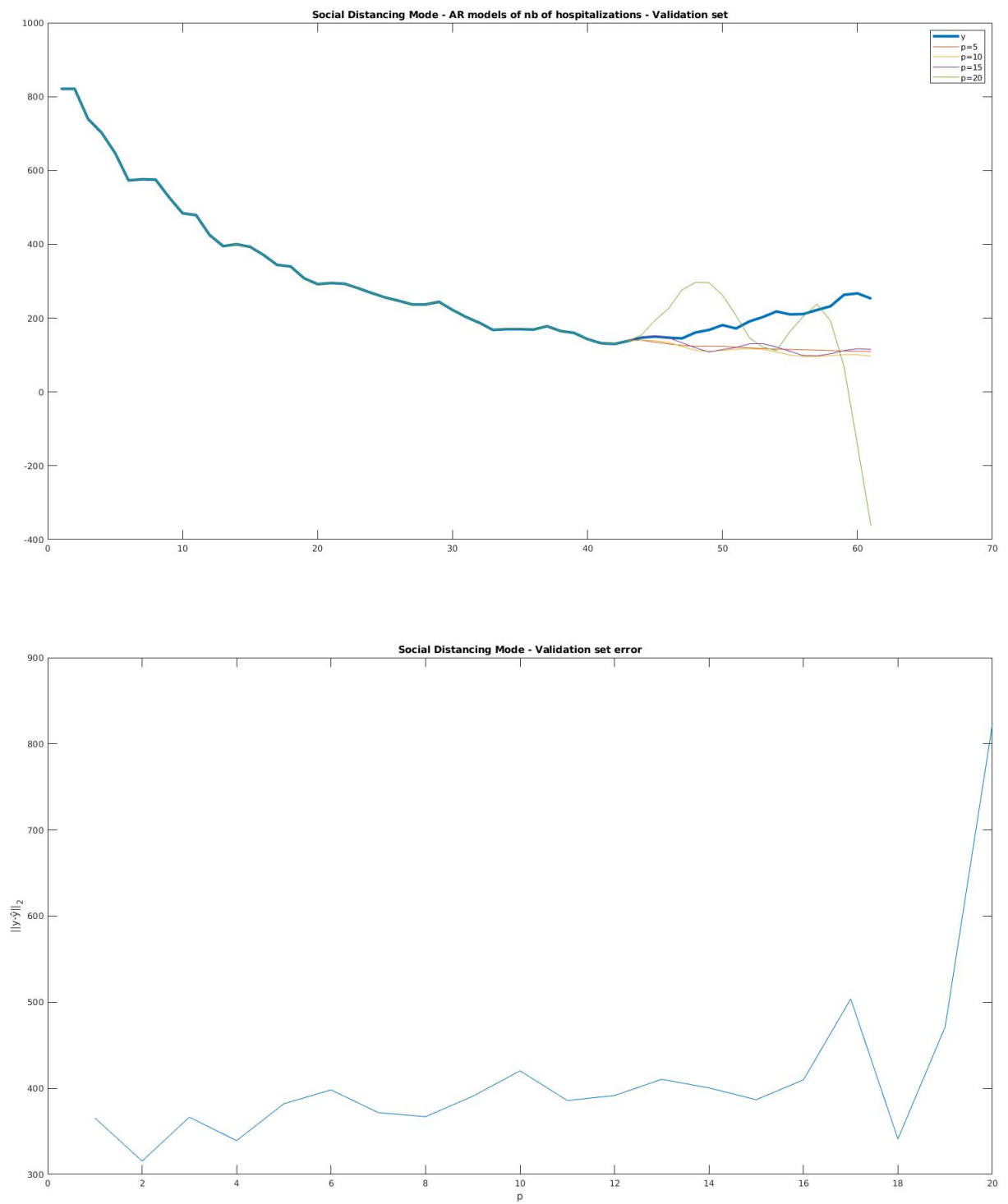
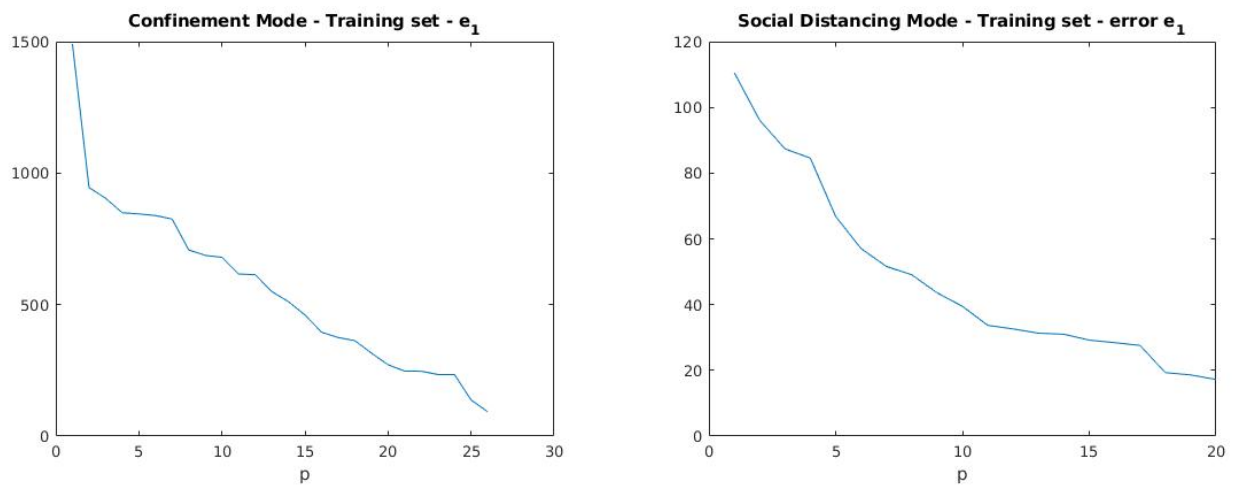Figure 4: Lockdown mode: validation set predictions and error using the AR model, for different values of p.

Figure 5: The error e_1 monotonically decreases.