Advanced Topics in Control 2020: Large-Scale Convex Optimization

Exercise 9: Regression

Goran Banjac, Mathias Hudoba de Badyn, Andrea Iannelli,
Angeliki Kamoutsi, Ilnura Usmanova

May 14, 2020

Date due: May 21, 2020 at 23:59.

Please submit your solutions via Moodle as a PDF with filename `Ex06_Surname.pdf`, replacing `Surname` with your surname.

---

1 Problem 1

Some of you may have heard of the Kalman filter, one of the greatest forms of wizardry known to the human race. In this exercise, you will derive the Kalman filter from a least-squares problem.

Recall some basic definitions from probability: for a random variable x, the mean and covariances are

$$\mathbb{E}[x] = \int x d\mu(x) \tag{1}$$

$$\mathrm{Cov}[x] = \mathbb{E}\left[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T\right]. \tag{2}$$

The measure of integration $\mu(x)$ is not important in this problem. Consider a linear system of the form

$$x_{k+1} = Ax_k + w_k \tag{3}$$

$$\tilde{y}_k = Cx_k + v_k, \tag{4}$$

where $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$ are zero-mean Gaussian random variables with covariances Q and R respectively.

(a) Suppose we don't have any information about the dynamics (3), i.e., we only have knowledge of the measurement model $\tilde{y}_k = Cx_k + v_k$. For now, let's (mostly) ignore the subscript k denoting the time index, since we don't know the dynamics. Show that the least-squares estimator in this case, given by

$$\hat{x} := \mathrm{argmin}_x \|\tilde{y} - Cx\|_2^2, \tag{5}$$

is *unbiased*, meaning that $\mathbb{E}[\hat{x}] = x$. For simplicity, assume that C has full column rank, so C^TC is invertible.

(b) Show that the *error covariance* P satisfies

$$P := \mathrm{Cov}[x - \hat{x}] = (C^T C)^{-1} C^T R C (C^T C)^{-1}. \tag{6}$$

(c) Suppose we have an unbiased estimate of x, which we denote $\hat{x}_1$, with error covariance P_1, and suppose we want to update it with some new measurement $\tilde{y}$. (This is referred to as updating an *a priori* estimate). We can solve the least squares problem

$$\hat{x}_2 = \mathrm{argmin}_x \left\{ (x - \hat{x}_1)^T P_1^{-1} (x - \hat{x}_1) + (\tilde{y} - Cx)^T (\tilde{y} - Cx) \right\}. \tag{7}$$

Show that $\hat{x}_2$ is an affine transformation of $\hat{x}_1$, i.e.

$$\hat{x}_2 = \Xi \hat{x}_1 + \beta, \tag{8}$$

by finding Ξ and β.

(d) In the least-squares problem (7), the first term of the objective is a *weighted least squares* function, where the weight is given by inverse of the *a priori* error covariance matrix P_1. Give an informal justification why doing this is a good idea.

(e) The objective function in (7) is not the only one we can use. Let's try to formulate a more general optimization problem to update the *a priori* estimate $\hat{x}^-$ with new data, but using the result of Part (c) as a stepping stone.

You should note that your result is a sum of two terms, corresponding to the old estimate and the new data, modulo some multiplication by matrices. Suppose our more general update takes the form

$$\hat{x}^+ = L\hat{x}^- + K\tilde{y}. \tag{9}$$

We want $\hat{x}^+$ to be unbiased. Using this assumption, show that $L + KC = I$, and that

$$\hat{x}^+ = \hat{x}^- + K(\tilde{y} - C\hat{x}^-). \tag{10}$$

(f) The objective function we want to minimize is actually the expectation of the least-squares error of $\hat{x}^+$, namely $\mathbb{E}\left[\|x - \hat{x}^+\|_2^2\right]$. Letting P^+ denote the error covariance $\mathrm{Cov}[x - \hat{x}^+]$ of $\hat{x}^+$, show that

$$\mathbb{E}\left[\|x - \hat{x}^+\|_2^2\right] = \mathrm{Tr}[P^+], \tag{11}$$

where $\mathrm{Tr}[A]$ refers to the *trace* of the matrix A, or the sum of the diagonal entries of A. If you are unfamiliar with the trace, the Wikipedia page on it lists some basic properties that you may need to solve this problem.

(g) We now want to find the optimal K in Equation (10) that minimizes (11). Using Equation (10), write $\mathrm{Tr}[P^+]$ in terms of K, and then by taking the first-order optimality condition

$$\frac{\partial}{\partial K} \mathrm{Tr}[P^+] = \mathbf{0}, \tag{12}$$

show that

$$K = P^- C^T \left(C P^{-1} C^T + R \right)^{-1}. \tag{13}$$

(h) By substituting your formula for K back into your equation for P^+ in terms of K from Part (g), show that

$$P^+ = (I - KC)P^-. \tag{14}$$

(i) We are almost done! So far, we have derived the *Kalman gain* in (13). The Kalman gain tells us how to optimally update our estimate given the difference between our estimate and the measurement we have received. Neat!

What is missing is how the dynamics plays into everything. Now, we will add our subscript k back into all of the quantities. The key thing to realize here is that the *updated estimate* $\hat{x}_k^+$ becomes our *a priori estimate* at the next timestep if we propagate it via the dynamics. Namely,

$$\hat{x}_{k+1}^- = A\hat{x}_k^+. \tag{15}$$

Using (15), show that

$$P_{k+1}^- = AP_k^+A^T + Q. \tag{16}$$

(j) To show me you understand what you just did, summarize the steps of one iteration of the Kalman filter algorithm you just derived.

2 Problem 2

Mathias, one of the postdocs at IfA, has gone insane due to the social distancing measures enforced by the Swiss government. On March 18th, he shaved his head. On March 30th, he was caught on Zoom speaking in a delicate voice to the eggplant[1] he is growing in a bucket in his home office. On April 27th, he transcended space and time, and now only speaks in modulated synthetic tones with frequencies given by his favourite sequence of numbers!

Desperate for help, and social contact, his ethereal corpse has broadcasted a distress signal, but 5G-tower saboteurs have caused the signal to be corrupted with noise! Oh no, what will we do?

His friends in the TA team of ATIC, who miss him terribly but mostly want him to help write the ATIC exam, are enlisting your help to de-code the message with your newly-obtained powers of least-squares estimation and signal de-noising.

Goran, an expert in Canadian dialects, has determined that Mathias speaks acausally by blurting out a tone with frequency content

$$f(t) = \sum_{i=1}^{N} a_i \cos(\theta_i t), \tag{17}$$

where a_i is the amplitude of the frequency θ_i. Cryptography experts at CSEC, the Canadian spy agency, have determined that:

- Each tone corresponds to a message
- Each frequency θ_i corresponds to a character (hence the word has N letters)
- The amplitude a_i is an integer from 1 to N corresponding to the position of the character.
- The frequencies θ_i are integers from Mathias' favourite sequences of numbers. Unfortunately, being a total nerd, he has many such sequences.

[1]The ATIC team has been informed that outside of Canada, this is referred to as an "aubergine".

(a) We will model the task of saving Mathias, like most important things in life, as a least-squares problem. First, we will define our model y and measurement $\tilde{y}$ as

$$y = Hx \tag{18}$$
$$\tilde{y} = Hx + v. \tag{19}$$

Here, $\tilde{y}$ represents our measurements corrupted by zero-mean Gaussian noise $v \sim \mathcal{N}(0, Q)$. What is the solution to the least-squares problem

$$\hat{x} := \operatorname{argmin}_x \|\tilde{y} - y\|_2^2 \tag{20}$$

in terms of H and $\tilde{y}$? Note that H will be chosen such that it has full column-rank.

(b) $\tilde{y}$ is a t_f-length vector containing the discretely-sampled components in time of the signal Mathias sent. In model form, we have

$$\begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_{t_f} \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\theta_1 t_1) & \cos(\theta_2 t_1) & \cdots & \cos(\theta_{27} t_1) \\ \cos(\theta_1 t_2) & \cos(\theta_2 t_2) & \cdots & \cos(\theta_{27} t_2) \\ \vdots & \vdots & & \vdots \\ \cos(\theta_1 t_f) & \cos(\theta_2 t_f) & \cdots & \cos(\theta_{27} t_f) \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} a_1 \\ \vdots \\ a_{27} \end{bmatrix}}_{x} + \begin{bmatrix} v_1 \\ \vdots \\ v_{t_f} \end{bmatrix}. \tag{21}$$

There are 26 letters of the alphabet plus the space character, and hence 27 possible frequencies in every signal.

In your favourite programming language `python`, write a function that returns such a matrix H for a given time horizon $t = [0, dt, 2dt, \ldots, (t_f - 1)dt]^T$ and set of frequencies $\theta_1, \ldots, \theta_{27}$. Note that H is simply,

$$H = \begin{bmatrix} | & & | \\ \cos(\theta_1 t) & \cdots & \cos(\theta_{27} t) \\ | & & | \end{bmatrix}. \tag{22}$$

(c) Assuming a time horizon t with $dt = 1/8192$, the default sampling rate in many audio applications, apply your estimator in part (a) using the H function from part (b) to the $\tilde{y}$ provided in `mathias_distress_call_1.csv`. The time horizon t_f is just the length of $\tilde{y}$. Use the set of frequencies generated by:

(i) The sequence defined by

"*T* is the first, fourth, eleventh, ... letter in this sentence,
not counting spaces or commas",

i.e. $n_1 = 1$, $n_2 = 4$, and so on. Mathias likes to start this sequence at the 25th entry, i.e. $\theta_1 = n_{25}$. (**Hint**: just google Aronson's sequence and use `ctrl+c/ctrl+v`).

(ii) Being disgusted with equal temperament, Mathias made up his own musical scale by dividing an octave into 27 equal notes. It is likely he would have broadcasted a message starting at E♭, the greatest of all notes, which has a frequency of 311.127. Hence,

$$\theta_k = 311.127 + \left(\frac{k-1}{27}\right) 311.127. \tag{23}$$

(iii) The "Fibonacci-on-steroids" sequence, with $\theta_1 = 150, \theta_2 = 175, m_1 = 0.5, m_2 = 0.8$:

$$\theta_k = \begin{cases} \theta_1 & k = 1 \\ \theta_2 & k = 2 \\ \lceil m_1 \theta_{k-1} \rceil + \lceil m_2 \theta_{k-2} \rceil & k \geq 3. \end{cases} \tag{24}$$

Which sequence of numbers gives you the most sensible solution, and for this sequence, what frequencies θ_i are present in the signal? What (integer) amplitudes a_i do these frequencies have? **Hint:** try rounding the solution, and checking its sparsity.

(d) What message has Mathias broadcasted? Assume that the frequencies $\theta_1, \ldots, \theta_{27}$ correspond to letters in order, with θ_{27} corresponding to a space:

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

(e) Mathias broadcasted a second message, but the noise is much larger due to the static electricity from the swarm of cats occupying his apartment. Using the same sets of frequencies from Part (c), use `cvxpy` to solve the LASSO problem

$$\min_x \|\tilde{y} - Hx\|_2^2 + \lambda\|x\|_1 \tag{25}$$

on the $\tilde{y}$ provided in `word2.csv`.

(i) What value of λ produces a good result? Give an informal justification of why the regularizer $\lambda\|x\|_1$ helps de-noise the signal.

(ii) Which one of Mathias' favourite sequences is used for this message?

(iii) What message has Mathias broadcasted? Does his message make sense given that he has shaved his head? Maybe it's for his cat...

(f) Notice that Mathias' method of encoding messages is pretty terrible. Due to aliasing issues (i.e., distinguishing each letter), in each message he can only send at most one of each letter and space. Is there a simple fix to this problem?