Assignment 5   2.086 Fall 2015

Due: *Tuesday, 27 October, at 5 PM.*

Upload your solution to Stellar as a zip file "`YOURNAME_ASSIGNMENT_5.zip`" which includes the script for each question in the proper format as well as any MATLAB functions (of your own creation) your scripts may call. Remember to verify your upload before the assignment due time by downloading your solution folder yourself and re-running `grade_o_matic`. Full instructions for assignment preparation are given in the document "Preparing assignments for `grade_o_matic`" available on the 2.086 website.

Questions

NOTE: In Q2, Q3, and Q4 we pose some optional questions related to real datasets. There is no direct credit associated with these optional questions, however, the real datasets do correspond to grader instances. We suggest you plot the data and your best-fit solution to confirm that your least-squares solution is reasonable.

1. (*20 points*) The stress-strain relation for an isotropic material in the elastic regime is given by Hooke's Law

$$\sigma^{\text{stress}} = E^{\text{true}} \varepsilon^{\text{strain}} , \tag{1}$$

where σ^{stress} is the uniaxial stress (in units of N/m^2), E^{true} is the (true) Young's modulus of the material (in N/m^2), and $\varepsilon^{\text{strain}}$ is the strain (dimensionless). We can write this more generally as a model for the stress (our dependent variable) in terms of the strain (our independent variable) and a parameter vector $\beta = (\beta_0 \;\; \beta_1)^{\text{T}}$:

$$\sigma^{\text{stress}}_{\text{model}}(\varepsilon^{\text{strain}}; \beta) = \beta_0 + \beta_1 \varepsilon^{\text{strain}} ; \tag{2}$$

it follows from Equation (1) that $\beta^{\text{true}} = (0 \;\; E^{\text{true}})^{\text{T}}$.

In order to estimate the parameter vector β^{true} (and hence the Young's modulus, E^{true}) we shall provide data: $m \times 1$ arrays `strains_meas` (the values of the strain at which we take stress measurements) and `stresses_meas` (the corresponding measured values of the stress). This data will of course reflect measurement noise. We shall estimate β^{true} by the least-squares solution $\hat{\beta}$. We recall that $\hat{\beta}$ minimizes

$$\left(\sum_{i=1}^{m} \left(\texttt{stresses\_meas(i)} - \sigma^{\text{stress}}_{\text{model}}(\texttt{strains\_meas(i)}; \beta) \right)^2 \right)^{1/2} , \tag{3}$$

over all possible 2-vectors β.

We obtain the least-squares solution $\hat{\beta}$ (`beta_hat` in MATLAB), for the particular case of m (MATLAB `m`) $= 8$ measurements, with the script provided immediately below.[1]

```
% begin script
```

[1]Note that prior to running the script the MATLAB workspace contains *only* `strains_meas` and `stresses_meas`.

```
m = 8;

X = [ones(m,1), EXPR1]; % EXPR1 to be specified in part (i)

beta_hat = X \ EXPR2;    % EXPR2 to be specified in part (ii)

% end script
```

(*i*) (5 points) In the script above, `EXPR1` should be taken as

(*a*) `stresses_meas`

(*b*) `strains_meas`

(*c*) `ones(m,1)`

(*d*) none of the above

(*ii*) (5 points) In the script above, `EXPR2` should be taken as

(*a*) `stresses_meas`

(*b*) `strains_meas`

(*c*) `ones(m,1)`

(*d*) none of the above

Note in the remainder of this question we assume that you have chosen the correct options above such that `beta_hat` of the script is equal to $\hat{\beta}$ which minimizes (3).

(*iii*) (5 points) Is it possible, *for some set of measurements*, that

$$\sigma^{\text{stress}}_{\text{model}}(\texttt{strains\_meas(i)}; \hat{\beta}) < \texttt{stresses\_meas(i)}$$

at all the data points, $1 \leq \texttt{i} \leq m$?

(*a*) Yes

(*b*) No

(*iv*) (5 points) Is it possible, *for some set of measurements*, that

$$\left(\sum_{i=1}^{m} \left(\texttt{stresses\_meas(i)} - \sigma^{\text{stress}}_{\text{model}}(\texttt{strains\_meas(i)}; \hat{\beta}) \right)^2 \right)^{1/2}$$

is *less than*

$$\left(\sum_{i=1}^{m} \left(\texttt{stresses\_meas(i)} - \sigma^{\text{stress}}_{\text{model}}(\texttt{strains\_meas(i)}; \beta^{\text{true}}) \right)^2 \right)^{1/2} ?$$

(a) Yes

(b) No

We recall that $\beta^{\text{true}} = \begin{pmatrix} 0 \\ E^{\text{true}} \end{pmatrix}$ and E^{true} is the true Young's modulus of the material.

The template `A5Q1_Template.m` contains the multiple-choice format required by `grade_o_matic`.

2. (*15 points*) A robot takes advantage of an IR Distance Transducer to identify the shape of a room and its own position within the room, as demonstrated in the 2.086 OCW video RoboRoom. In this question we would like you to develop a script which, given a set of measured (transducer voltage, distance) pairs, provides estimates for the calibration constants which inform the transducer voltage $(V) \to$ distance (D) relationship exploited by the robot.

   We shall consider for our transducer voltage $(V) \to$ distance (D) relationship a power law $D^{\text{model}}(V) = CV^{-\gamma}$ characterized by the two calibration parameters, C and γ. We then follow the "log transformation" procedure of page 47 of the Cumulative Class notes to deduce estimates $\hat{C}$ and $\hat{\gamma}$ for the calibration parameters C^{true} and γ^{true}, respectively. Two cautionary notes: the "log" in the "log transformation" is the naural logarithm (base e); the measurements, which shall be inputs to your procedure, are provided as (voltage, distance) pairs, *not* (log(voltage), log(distance)) pairs.

   Your script takes two script inputs, V_i and D_i^{meas}, for $1 \leq i \leq m$: D_i^{meas} is an independently measured distance corresponding to the transducer voltage V_i. These inputs must correspond in your script to MATLAB $m \times 1$ arrays `V` and `D_meas`: entry i of `V` and `D_meas` contains, respectively, the voltage V_i (in volts) and the distance D_i^{meas} (in cm) associated with the i^{th} calibration pair. Allowable instances must satisfy $1 \leq m \leq 10000$ and furthermore yield a design matrix X with independent columns. (Note that m is not a script input and should instead be deduced from (say) `D_meas`.)

   The script yields two script outputs: $\hat{C}$ and $\hat{\gamma}$, your estimates (as defined above) for the calibration constants C^{true} and γ^{true}, respectively; $\hat{C}$ and $\hat{\gamma}$ must correspond in your script to MATLAB scalars `C_hat` and `gamma_hat`, respectively.

   A template is provided in `A5Q2_Template`. We emphasize that your script should perform correctly for any set of (real or synthetic) data. You should yourself devise several test cases for which you can anticipate the correct answers and hence test your script.

   Begin Optional. We do provide you with one real dataset if you would like to exercise your script in an actual engineering context. The mysterious Dr James Penn conducted calibration experiments for a particular transducer, a Sharp GP2Y0A02YK0F (with a 15-150 cm range). The experimental data comprises $m = 15$ measurements — calibration pairs — provided to you in the `.mat` file `IR_Transducer_Data` (available in the `Assignment_5_Templates` folder) as $m \times 1$ arrays `V` and `D_meas`; the latter conform to the script input specifications. Based on your analysis, and graphical inspection, do you believe that the resulting calibration curve — $\hat{D}(V) = \hat{C}V^{-\hat{\gamma}}$ — will be uniformly accurate over the entire range of distances

15 cm $\leq D \leq$ 150 cm? Do you find any evidence for model error in the assumed form of the calibration curve, $D^{\text{model}}(V) = CV^{-\gamma}$? *End Optional*

3. (*15 points*) As demonstrated in the 2.086 OCW video RoboFriction, the friction coefficient between a robot wheel and the ground plays a crucial role in robot navigation and performance. We would like to predict the friction force for the materials relevant to this particular robotic application. Towards that end, we would like you to develop estimates $\hat{C}$, $\hat{\alpha}$, and $\hat{\eta}$ for the parameters C^{true}, α^{true}, and η^{true}, respectively, associated with the power law friction model proposed in **CYAWTP 50** of the 2.086 Cumulative Class Notes, Chapter 6.

   We ask that you apply the "log transformation" approach: transform the power friction law to a linear form in three parameters, β_0, β_1, and β_2; apply least squares to construct estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$, for β_0, β_1, and β_2, respectively; infer from (your log transformation, and) $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ your estimates $\hat{C}$, $\hat{\alpha}$, and $\hat{\eta}$ for C, α, and η, respectively. You are provided with m experimental measurements of the form $((F_{\text{normal}}, A_{\text{surface}})_i, (F_{\text{f, static}}^{\text{max, meas}})_i), 1 \leq i \leq m$; here $(F_{\text{f, static}}^{\text{max, meas}})_i$ is the measurement of the maximum static friction force corresponding to the (normal force, superficial surface area) pair $(F_{\text{normal}}, A_{\text{surface}})_i$.

   The script takes three script inputs: $(F_{\text{normal}})_i$, $(A_{\text{surface}})_i$, and $(F_{\text{f, static}}^{\text{max, meas}})_i$, for $1 \leq i \leq m$, where $(F_{\text{f, static}}^{\text{max, meas}})_i$ is the maximum measured static friction force corresponding to the normal load $(F_{\text{normal}})_i$ and the superficial surface area $(A_{\text{surface}})_i$. These inputs must correspond in your script to MATLAB $m \times 1$ vectors `F_normal`, `A_surface`, and `F_fstaticmaxmeas`: the entry i of `F_normal`, `A_surface`, and `F_fstaticmaxmeas` contains, respectively, the prescribed normal load F_{normal} (in Newtons), the prescribed surface area A_{surface} (in cm^2), and the measured friction force $F_{\text{f, static}}^{\text{max, meas}}$ (in Newtons) for the i^{th} measurement. Allowable instances must satisfy $1 \leq m \leq 10000$ and furthermore yield a design matrix X with independent columns. (Note that m is not a script input and should instead be deduced from (say) `F_fstaticmaxmeas`.)

   The script yields three script outputs: scalars $\hat{C}$, $\hat{\alpha}$, and $\hat{\eta}$, which must correspond in your script to MATLAB variables `C_hat`, `alpha_hat`, and `eta_hat`, respectively.

   A template is provided in `A5Q3_Template`. We emphasize that your script should perform correctly for any set of (real or synthetic) data. You should devise several test cases for which you can anticipate the correct answers and hence test your script.

   Begin Optional. We do provide you with a real (noisy) dataset in the event that you would like to exercise your script in an actual physical context. The indefatigable Dr James Penn conducted experiments, for a particular pair of robot-wheel and ground materials, to obtain friction force measurements $F_{\text{f, static}}^{\text{max, meas}}$ (in Newtons) as a function of normal load F_{normal} (in Newtons) and superficial contact area A_{surface} (in cm^2). The turntable apparatus is shown in Figure 1.

   The experimental data comprises $m = 50$ measurements: 2 (repetition) measurements at each of 25 points on a 5×5 "grid" in $(F_{\text{normal}}, A_{\text{surface}})$ space. This *real* data is provided to you in the `.mat` file `friction_data` (available in the `Assignment_5_Templates` folder) as 50×1 arrays `F_normal`, `A_surface`, and `F_fstaticmaxmeas`: entry i of `F_normal`, `A_surface`, and `F_fstaticmaxmeas` provides, respectively, the prescribed normal load F_{normal} (in Newtons), the prescribed surface area A_{surface} (in cm^2), and the measured friction force $F_{\text{f, static}}^{\text{max, meas}}$ (in Newtons) associated with the i^{th} measurement; for example, in the first measurement, $i = 1$, the imposed normal load is 0.9810 Newtons, the superficial contact area is 1.2903 cm^2, and the resulting measured friction force is 0.1080 Newtons.

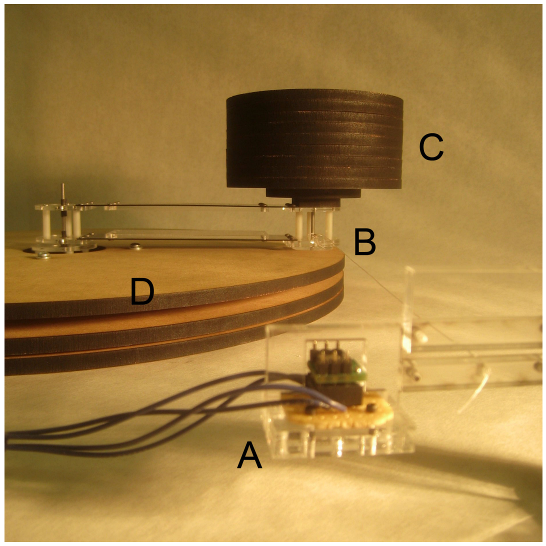

Figure 1: Experimental apparatus for friction measurement: Force transducer (A) is connected to contact area (B) by a thin wire. Normal force is exerted on the contact area by load stack (C). Tangential force is applied on turntable (D) and transmitted by friction from the turntable surface to the contact area. Apparatus and photograph courtesy of James Penn.

You may recall Amontons' "Law," which states that the friction force is linearly dependent on normal load — our constant C is the static friction coefficient, μ — and independent of superficial contact area. Based on Dr Penn's data, did Amontons get it right, at least for the particular materials of our 2.086 experiment? In other words, would you conclude, for our pair of materials, that the static friction force is indeed independent of the superficial contact area? (A more rigorous analysis of this conjecture would require suitable assumptions on the measurement noise on the basis of which we could then construct confidence intervals for our parameters.) *End Optional*

4. *(15 points)* In 2.086, in the Spring of 2013 (during which the ubiquitous Dr Penn served as an Instructor), a quiz — the same quiz — is administered asynchronously across recitations which take place on the Monday, Wednesday, and Friday of a particular week. We would like to understand if, on average, the students who take the quiz later in the week perform the same as the students who take the quiz earlier in the week. In particular, we do not wish to afford any groups of students an unfair advantage.

   We first postulate a model for the average grade received on the quiz as a function of time,

$$g^{\text{model}}(t;\beta) = \beta_0 + \beta_1 t \ , \tag{4}$$

   where g represents the average numerical score (out of 100 points), $\beta = (\beta_0 \ \beta_1)^{\mathrm{T}}$ are the coefficients to be determined from data, and t is the time at which the quiz is administered. We shall measure time in days such that the Monday, Wednesday, and Friday of the quiz week correspond to $t = 1$, $t = 3$, and $t = 5$, respectively.

   We next assume that the scores for the m students who shall take the quiz can be represented as

$$g_i^{\text{meas}} = g^{\text{model}}(t_i;\beta^{\text{true}}) + \epsilon_i \ , \tag{5}$$

where g_i^{meas} is the grade received by student i, β^{true} is the true value of the parameter vector β in (4),[2], t_i is the time — day 1, 3, or 5 — at which student i takes the quiz, and ϵ_i reflects variations in performance amongst individual students in the class.

In this question we would like you to write a script which, for some given set of student grades, performs a least-squares fit to identify an estimate $\hat{\beta}_1$ for β_1^{true}, the "slope" of our grade model (4).

The script takes three script inputs: a $m_1 \times 1$ array of the grades of the students who take the quiz on Monday (hence the g_i^{meas} for all students i for whom $t_i = 1$), which must correspond in your script to MATLAB variable `Mon_grades`; a $m_3 \times 1$ array of the grades of the students who take the quiz on Wednesday (hence the g_i^{meas} for all students i for whom $t_i = 3$), which must correspond in your script to MATLAB variable `Wed_grades`; and a $m_5 \times 1$ array of the grades of all students who take the quiz on Friday (hence the g_i^{meas} for all students i for whom $t_i = 5$), which must correspond to MATLAB variable `Fri_grades`. Allowable instances of each of these three arrays must satisfy $0 \leq g. \leq 100$ and furthermore $m_1 \geq 1, m_3 \geq 1$, and $m_5 \geq 1$. (Note that m_1, m_3, and m_5 are not inputs: these variables should instead be deduced from `Mon_grades`, `Wed_grades`, and `Fri_grades`, respectively.)

The script yields a single output: $\hat{\beta}_1$, which must correspond in your MATLAB script to the scalar `beta_hat_1`.

A template is provided in `A5Q4_Template`. We emphasize that your script should perform correctly for any set of (real or synthetic) data. You should devise several test cases for which you can anticipate the correct answers and hence test your script.

Begin Optional. We provide you with the real data of Spring 2013 in the event that you wish to exercise your script in an actual academic context. The experimental data of the asynchronous quiz "experiment" of Spring of 2013 is summarized in the files `Mon_grades`, `Wed_grades`, and `Fri_grades` in the `.mat` file `Quiz_1_S2013_Grades` (available in the `Assignment_5_Templates` folder); these three files conform to the script input prescriptions for `A5Q4.m`. Based on this data, would you conclude that the students who take the quiz later in the week perform the same as the students who take the quiz earlier in the week? (A more rigorous analysis of this hypothesis would require suitable assumptions on the distribution of grades (over the student population) on the basis of which we could then construct a confidence interval for β_1^{true}.)
End Optional

5. (*20 points*) The height of a ball as a function of time is given by $z^{\text{true}}(t)$. We wish to determine the acceleration of the ball, $\ddot{z}(t)$, from noisy measurements. Note that z is measured positive upwards relative to the acceleration of gravity.

   In particular, for any given time t_0 of interest, we are provided with $m = 5$ measurements, $(t_{-2}, z_{-2}^{\text{meas}}), (t_{-1}, z_{-1}^{\text{meas}}), (t_0, z_0^{\text{meas}}), (t_1, z_1^{\text{meas}}), (t_2, z_2^{\text{meas}})$. The measurement times $t_i, -2 \leq i \leq 2$, are equispaced:

$$t_i = t_0 + i\,\delta t\ , -2 \leq i \leq 2\ , \tag{6}$$

   for a given positive δt.

[2] We may think of β^{true} as the value of the parameter vector in the hypothetical limit of an infinite number of students.

We consider two approximations to $\ddot{z}(t_0)$, the first based on finite-difference approximation, the second based on least-squares fit. For the finite-difference approximation, we take $\ddot{z}(t_0) \approx \ddot{z}_{\delta t}(t_0)$, where

$$\ddot{z}_{\delta t}(t_0) \equiv \frac{z_{-1}^{\text{meas}} - 2z_0^{\text{meas}} + z_1^{\text{meas}}}{(\delta t)^2} \ . \tag{7}$$

For the regression approximation, we take $\ddot{z}(t_0) \approx \ddot{\hat{z}}(t_0)$, where $\hat{z}(t) = z^{\text{model}}(t;\hat{\beta})$ is provided by the least-squares fit of the data (t_i, z_i^{meas}), $-2 \le i \le 2$, to the quadratic model $z^{\text{model}}(t;\beta) = \beta_0 + \beta_1 t + \beta_2 t^2$. The latter is, of course, motivated by Newton's Laws for the motion of a body in a vacuum acted upon by a constant gravitational field.

In this question we would like you to write a script which, for some given set of height measurements,

$$(t_{-2}, z_{-2}^{\text{meas}}), (t_{-1}, z_{-1}^{\text{meas}}), (t_0, z_0^{\text{meas}}), (t_1, z_1^{\text{meas}}), (t_2, z_2^{\text{meas}}),$$

evaluates $\ddot{z}_{\delta t}(t_0)$ and $\ddot{\hat{z}}(t_0)$.

The script takes two script inputs: t_i, $-2 \le i \le 2$, and z_i^{meas}, $-2 \le i \le 2$, where z_i^{meas} is the measured height (in meters) corresponding to the time t_i (in seconds). These inputs must correspond in your script to MATLAB 5×1 vectors `time_vec` and `z_meas_vec`, respectively: for $k = 1, \ldots, 5$, entry k of `time_vec` (respectively, `z_meas_vec`) is t_{k-3} (respectively, z_{k-3}^{meas}). (Note we shift the indices such that the MATLAB array first index is always 1.) Allowable instances must conform to (6), and yield a design matrix X with independent columns.

The script yields two script outputs: acceleration estimates $\ddot{z}_{\delta t}(t_0)$ and $\ddot{\hat{z}}(t_0)$, which must correspond in your function to MATLAB scalar variables `accel_fd` and `accel_fit`, respectively.

A template is provided in `A5Q5_Template`. We emphasize that your script should perform correctly for any set of (real or synthetic) data. You should yourself devise several test cases for which you can anticipate the correct answers and hence test your script.

6. (*15 points*) In this question we provide you with a real (noisy) dataset for the falling ball experiment of Question 5. The illustrious Dr James Penn conducted falling ball experiments for a ball of mass 0.014 kg and diameter 0.0508 meters: a video camera records the trajectory; an image processing technique, with correction for lens distortion, yields data for height as a function of time. We provide to you in the `.mat` file `Falling_Ball_Snippet` (available in the `Assignment_5_Templates` folder) as 5×1 arrays `time_vec` and `z_meas_vec`. The `time_vec` and `z_meas_vec` arrays conform to (6) and yield a design matrix X with independent columns, and thus are allowable input instances for your script. Note that in Dr Penn's data, height is defined as distance above the ground — hence positive, and decreasing as a function of time.

   Exercise your script of Question 5 to develop finite-difference and best-fit estimates for the acceleration at time t_0, $\ddot{z}_{\delta t}(t_0)$ and $\ddot{\hat{z}}(t_0)$, respectively. Which approximation, finite-difference (based on interpolation) or best-fit, is most accurate? Justify your answer both in terms of physical arguments — anticipated behavior — and also mathematical arguments — the effect of noise. Also provide a plot, with axes time and height, which includes (a) the data, (t_i, z_i^{meas}), $-2 \le i \le 2$, (b) the best-fit solution $\hat{z}(t)$ (for some suitably fine grid of time points over the interval $[t_{-2}, t_2]$), and (c) the quadratic interpolant through t_{-1}, t_0, and t_1 (see page 10 of the Cumulative Class Notes) on which the finite-difference approximation is based; make sure to label axes, include a legend, and provide a figure title.

Please provide your answer in the form of a .pdf document `A5Q6.pdf` which you should include in your folder `YOURNAME_ASSIGNMENT_5.zip` which you upload to Stellar.