Project title: Milestone 1 - Predicting seismic collapse capacity of a 4-story steel structure

Short problem description

Earthquakes are seismological events that cause the ground to start shaking. The primary interest of structural earthquake engineers is not on the earthquake per se, but rather on the resulting ground shaking beneath a structure of interest, for instance an existing building or a planned design. This is because this shaking of the ground, which is referred to as an 'earthquake ground motion', imposes dynamic loads on the structures. Depending on the intensity of the ground motion, the imposed loads can be larger or smaller and it is the responsibility of engineers to ensure the structure achieves adequate and desired performance given the seismological setting.

One of the important ways in which structural performance can be measured relates to the ability of a structure to withstand earthquake loads without collapsing. It is intuitively clear that in order to induce collapse of a structure, a ground motion needs to have enough intensity (or, conversely, that the structure is 'weak' compared to the ground motion). This intensity at which a ground motion will induce collapse of a structure is called collapse capacity of a ground motion. There are formal ways how this value can be computed and you will learn about those in your future classes. Suffice it to say that this computation is numerically very expensive, i.e. it takes a long time even if performed on a cluster. So, from a practical viewpoint, it would be beneficial to develop a machine learning or statistical model to efficiently estimate the collapse capacity without having to perform expensive numerical computations on the cluster. This is where you come in!

The friendly people from RESSLab@EPFL have pre-computed collapse capacities of a 4-story steel structure for a large database of ground motions from past earthquakes around the world. Given this data your task is to develop an ML tool that predicts collapse capacities for future earthquakes. While you will be working with a single structure, there are many potential applications for this tool. For instance, when an earthquake occurs, such a tool could be used for rapid assessment to gauge which of the buildings have likely collapsed allowing for a more focused emergency response. This is just one application example, but if you are creative this stuff really has startup potential — google 'One Concern' for an example.

Brief background info on structural dynamics

The majority of problems you encountered so far in structural analysis probably involved the so called static loads, i.e. the types of loads that do not change as a function of time. For instance, self-weight of a building or the occupant loads can be appropriately approximated as static loads. Earthquake loads on the other hand are dynamic loads that significantly depend on time. They typically last a short time, but the shaking can be quite intense. As such they induce a dynamic response of a structure, e.g. a building starts shaking as a result of the imposed ground motion. Much like how different people prefer different types of music, and hence dance differently to different songs, so do different buildings respond differently to a given ground motion. How a particular building responds to a given earthquake loading, depends on the dynamic properties of that building.

A basic dynamic property of a building is the fundamental period of the building, often indicated as T_1. This is essentially the same characteristic as a period of a pendulum you learned about in physics class — it describes the time it takes a pendulum to complete one cycle of free oscillation. Specifically, if you displace a pendulum and let it go, the time it takes the pendulum to return to the starting position is the period of that pendulum. Similarly, if you were to pull the roof of a building and let it go, the time it would take the building to do one cycle of oscillation and return to the starting position is the period of the building. It turns out that this fundamental period T_1 is closely related to familiar concepts of stiffness and mass. You can learn more about these things in structural dynamics and earthquake engineering classes.

A brief and gentle introduction to earthquake ground motion intensity measures (IMs)

As mentioned, earthquakes are seismological events that cause the ground to start shaking. Earthquakes occur on the so-called faults which are fractures within the Earth's crust along which the parts of the crust on the opposite sides of the fault move relative to each other. As a result of this relative movement, or slip on the fault, the earth surrounding the fault starts to shake. This shaking propagates through the ground in form of seismic waves which reach different points on the Earth.

Intuitively it is clear that depending on where you are located in relation to the origin of the earthquake, the ground shaking you experience will differ. For instance, if an earthquake happens on the northern portion of the San Andreas fault in California, then people in San Francisco (West Coast of the US) will experience much stronger shaking than people in New York (East Coast of the US)[1] or Lausanne. But the differences in shaking are felt on a much smaller scale than that – even people, and likewise buildings, in different parts of San Francisco will experience different levels and types of shaking. Since it is the shaking of the ground that induces dynamic response of buildings and other structures, it is important to differentiate between an earthquake as seismological event and the shaking of the ground, or the so-called ground motions resulting from the earthquake. Important thing to note here is that for a single earthquake there are many resulting ground motions.

Ground motions are recorded as time series representing the acceleration of the ground shaking as a function of time, which is called a seismogram and is typically indicated as a_g (standing for acceleration of the ground). Shown in Figure 1 is the seismogram of a ground motion recorded during the March 22, 2020 Zagreb, Croatia earthquake. As the movement of the ground is three-dimensional, the seismograms are typically recorded in three directions and represented by two orthogonal horizontal components and one vertical component. Shown in Figure 1 is the north-south component of the ground motion recorded

[1] This is an excellent place to make a plug for the severely scientifically flawed, but nevertheless best movie ever – San Andreas the movie: 'You will feel it on the East Coast!'

in Zagreb and the units of the seismogram are in terms of the gravitational acceleration, or *g*, which equals 9.81 m/s^2.

The intensity of earthquakes, as opposed to ground motions, is typically described in terms of a magnitude which is a measure of the energy released when the fault ruptures. For instance, the magnitude of the earthquake that caused the ground motion depicted in Figure 1 was 5.3 on the Richter scale. This is a magnitude associated with events that occur in regions of low to moderate seismicity, for instance in Croatia or Switzerland. On the other hand, events on the before mentioned infamous San Andreas fault can be much larger, such as the 1906 San Francisco earthquake with magnitude 7.9. In a different seismological setting, specifically in the so-called subduction zones, earthquakes can be stronger still. This was the case with the 2011 Tohoku, Japan earthquake (magnitude 9.1) which in addition to strong shaking of the ground also triggered a huge tsunami which caused widespread destruction including a nuclear accident in the Fukushima Daiichi Nuclear Power Plant.

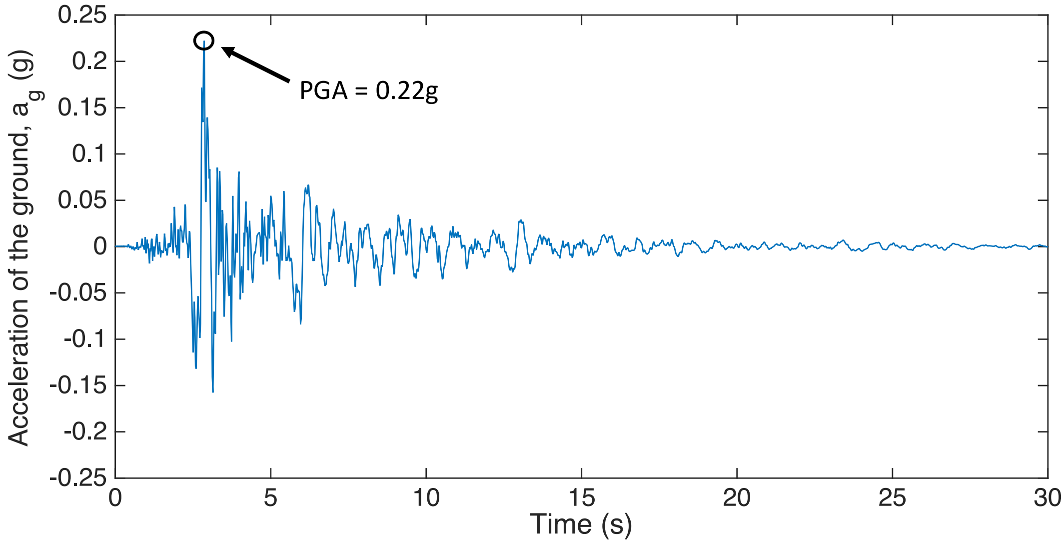

Figure 1. The north-south component of a seismogram recorded during the March 22, 2020, earthquake in Zagreb, Croatia.

While the magnitude of the causal earthquake is related to the properties of resulting ground motions, perhaps it should not be surprising that a single scalar that describes the energy released during the earthquake would not be a very good descriptor of complex waveforms describing the shaking of the ground at different locations. Since structural engineers are interested in dynamic responses of structures due to ground shaking, it is necessary to introduce additional metrics that more closely describe the intensity of the ground motions (rather than describing the earthquake). These metrics are called ground motion intensity measures (IMs) and one example of an IM is the peak ground acceleration, or PGA. As the name implies, PGA describes the absolute peak acceleration observed in a seismogram. For the seismogram shown in Figure 1, the value of PGA equals 0.22*g* and its occurrence in the waveform is indicated with a small circle. This value means that the structure experiences 22%, or about one fifth of the gravitational acceleration during the shaking in the horizontal direction and this acceleration is introduced into the structure at its base. For comparison, the largest PGA recorded during the 2011 Tohoku earthquake was

2.82*g* or almost three times the gravitational acceleration. To put things in context and help with intuitive understanding of these values, imagine that you rotate the building for 90 degrees such that it lies sideways. In this case, the building would need to withstand gravitational acceleration of 1.0*g*. It should be clear from this discussion that the earthquake ground motions can be very intense and that civil engineers have an important societal responsibility in ensuring resiliency of our infrastructure.

One of the limitations of PGA as an intensity measure is that it is computed solely from the ground motion and it does not take into account any dynamic property of a structure that may be exposed to that ground motion. For instance, it turns out that a ground motion with a large PGA can severely affect a nuclear power plant while a tall building may not be affected at all. What is at play here is that different structures have different dynamic characteristics, primarily their periods of vibration, so they are differently affected even by the same ground motion. To address this limitation of PGA, researchers have come up with a concept of the so called spectral acceleration which is indicated as *Sa(T)*. Loosely speaking, these spectral accelerations can be thought of as an extension of the PGA concept in a way that takes into account the period T of the case study structure. In other words, they describe accelerations that structures with different periods will experience when excited by the same ground motion (note: this is a very informal description and you can learn about these things in structural dynamics and earthquake engineering classes). The dataset provided with this project contains values of spectral accelerations, *Sa(T)*, computed at a range of periods spanning different structures that may be of interest. For instance, nuclear power plants would have short periods (e.g. fundamental period T_1 = 0.05 seconds) while tall buildings would have much longer periods (e.g. T_1 = 4s) so the difference in the effect of the same ground motion on those structures would be reflected in different values of spectral accelerations, namely *Sa(T=0.1s)* as opposed to *Sa(T=4s)*. In the context of this project, the collapse capacity of a structure is typically expressed as the value of the smallest spectral acceleration of a ground motion at the fundamental period of the structure which will cause the structure to collapse. The collapse capacity is typically indicated as $Sa(T_1)$@collapse.

The last intensity measure we will mention here pertains to the duration of the ground motion. Intuitively it makes sense that the longer the ground shaking lasts, the more damaging the ground motion could be for the structure. This is especially true when considering if a ground motion will induce a collapse of a structure. There are different ways how one may quantify the duration of a ground motion, but a key characteristic to capture is the portion of a ground motion that causes significant response of a structure. For instance, the length of the seismogram in Figure 1 is about 30 seconds, but the significant duration of this record only lasts about 3 to 4 seconds. The values of significant durations of ground motions that you will use for the project are also precomputed and provided with the dataset.

Description of the dataset

You are provided with files containing metadata on ground motion intensity measures for about 17,000 past earthquakes recorded around the world. Specifically, the following ground motion intensity measures are included in the dataset:

- Spectral accelerations, *Sa(T)*, computed at 105 different periods ranging from 0.01s up to 10s. These values are in columns with names representing the period in seconds, while the values in the rows represent spectral accelerations in units of 'g' (i.e. gravitational acceleration).
- Average spectral accelerations, Sa_{avg}, in units of g. As the name implies, this IM was computed by averaging the *Sa(T)* values over an appropriately selected period range (averaging period range depends on the structure).
- Two different measures of ground motion durations (da5_75 and da5_95) in units of seconds.
- Filtered incremental velocity, FIV3 (units: cm/s). This is an interesting IM, but we did not mention it in the *brief* intro to IMs as a guy at Stanford devoted his PhD life to developing this metric (i.e. it is a bit involved to explain). For his sake, hopefully you will find this IM to be a good predictor!

In addition to intensity measures of the ground motions, the provided files also contain the collapse capacities (sat1_col) of a 4-story steel structure computed for those ground motions. These collapse capacities were computed using the numerical models of the building and the computations were performed on the FIDIS cluster at EPFL.

Project description tl;dr

You are provided with a set of predictive features (earthquake ground motion intensity measures, IMs) and the associated responses (collapse capacities of a 4-story steel structure). Be creative and figure out the relationship between these values so as to be able to predict the response given previously unseen inputs.

Problem set questions

- What is your suggested way of addressing this problem? That is, which type of an approach would you use?
- What do the features of the dataset look like? How can you pre-process this data?
- Do the trends in the data make sense in your opinion and given your current understanding of the effect of earthquakes on the structures? For instance, what does your intuition tell you the effect of ground motion duration (i.e. how long the earthquake shaking lasts) would be on the collapse capacity of the ground motion?
- Which machine learning techniques did you try? Which of them worked the best, and why do you think this is the case?