Solution to Test 1

Problem 1. A function $f(x)$ is called *shift-scale-invariant* if for every x_0, there exists a value μ such that $y = f(x)$ implies $y' = f(x')$, where we denoted $y' = \mu \cdot y$ and $x' = x + x_0$. Prove that every differentiable shift-scale invariant function has the form $f(x) = A \cdot \exp(a \cdot x)$ for some A and a.

Proof. Let us assume that the differentiable function $f(x)$ is shift-scale-invariant. By definition of shift-scale-invariance, this means that for every x_0, there exists some value μ (depending on x_0) for which $y = f(x)$ implies that $y' = f(x')$, where $y' = \mu \cdot y$ and $x' = x + x_0$. Since μ depends in x_0, let us write this dependence in explicit form $\mu = \mu(x_0)$.

Let us take any x and take $y = f(x)$. Then, for each x_0, we have $y' = f(x')$, where $y' = \mu(x_0) \cdot y$ and $x' = x + x_0$. Substituting these expressions for y' and x' into the formula $y' = f(x')$, we conclude that

$$\mu(x_0) \cdot y = f(x + x_0).$$

Here, by our choice of y, we have $y = f(x)$. Substituting $f(x)$ instead of y into the above equality, we get

$$\mu(x_0) \cdot f(x) = f(x + x_0).$$

Let us now differentiate both sides of this equality with respect to x_0: since the functions are equal, their derivatives should be equal too.

With respect to x_0, the term $f(x)$ – that does not depend on x_0 – is a constant. Thus, the derivative of the left-hand side takes the form

$$\frac{d\mu}{dx_0} \cdot f(x).$$

To compute the derivative of the right-hand side, we use the chain rule, and get

$$\frac{d}{dx_0} f(x + x_0) = \frac{df}{dx}(x + x_0) \cdot \frac{d}{dx_0}(x + x_0) = \frac{df}{dx}(x + x_0).$$

Thus, the equality between derivatives of the left-hand side and of the right-hand side takes the form

$$\frac{d\mu}{dx_0} \cdot f(x) = \frac{df}{dx}(x + x_0).$$

This equality is true for every x_0. To simplify this equality, let us take $x_0 = 0$, and let us denote by a the value of the derivative $\dfrac{d\mu}{dx_0}$ for $\lambda = 1$. Then, we get the following:

$$a \cdot f = \frac{df}{dx}.$$

Let us separate the variables. For this, we divide both sides of this equality by f and multiply both sides by dx. Then, we get:

$$a \cdot dx = \frac{df}{f}.$$

Now, we integrate both sides, and get

$$a \cdot x + C = \ln(f),$$

where C is the integration constant. By applying $\exp(x)$ to both sides of this equality and taking into account that $\exp(\ln(x)) = x$ – by definition of the logarithm – we get

$$f(x) = \exp(a \cdot x + C).$$

We know that for all A and B, we have $\exp(A + B) = \exp(A) \cdot \exp(B)$. Thus,

$$f(x) = \exp(C) \cdot \exp(a \cdot x),$$

i.e., the desired form $f(x) = A \cdot \exp(a \cdot x)$, where we denoted $A \stackrel{\text{def}}{=} \exp(C)$.

The statement is proven.

Problem 2. Suppose that you know the values of some quantity v in two points x_1 and x_2, these values are $v_1 = 100$ and $v_2 = 200$. Based on this information, we want to use the inverse distance weighting technique to predict the value v of this quantity as a point x for which $d(x_1, x) = 10$ and $d(x, x_2) = 20$. Take $a = -1$.

Reminder. The general formula has the form

$$v = \frac{\sum\limits_i v_i \cdot (d(x, x_i))^a}{\sum\limits_i (d(x, x_i))^a}.$$

Solution. In our case, the numerator is equal to

$$v_1 \cdot d(x, x_1)^{-1} + v_2 \cdot d(x, x_2)^{-1} = \frac{100}{10} + \frac{200}{20} = 10 + 10 = 20,$$

and the denominator is equal to

$$d(x, x_1)^{-1} + d(x, x_2)^{-1} = \frac{1}{10} + \frac{1}{20} = \frac{2}{20} + \frac{1}{20} = \frac{3}{20}.$$

Thus, the desired value is equal to

$$v = \frac{20}{\frac{3}{20}} = \frac{20 \cdot 20}{3} = \frac{400}{3} = 133.3\ldots$$

Problem 3. Use calculus to find the value x for which the following function attains is minimum $2x^2 - 3x + 1$. What is the value of this minimum?

Solution. According to calculus, minimum and maximum of a function are attained at the points at which its derivative is equal to 0. Here,

$$(2x^2 - 3x + 1)' = 4x - 3,$$

so the location of the minimum can be determined by the equation $4x - 3 = 0$, so $4x = 3$, and $x = \frac{3}{4}$.

For this x, the function takes the value

$$2x^2 - 3x + 1 = 2 \cdot \left(\frac{3}{4}\right)^2 - 3 \cdot \frac{3}{4} + 1 = 2 \cdot \frac{9}{16} - \frac{9}{4} + 1 = \frac{9}{8} - \frac{9}{4} + 1 = \frac{9}{8} - \frac{18}{8} + \frac{8}{8} = -\frac{1}{8}.$$

Problem 4. Describe a function $y = 1/(1+x^2)$ as a composition of invariant functions.

Comment. This function is actively used in physics and in uncertainty quantification.

Solution. How do we compute this function? First, we compute the first intermediate value

$$z_1 = x^2.$$

In terms of z_1, the desired expression takes the form $y = 1/(1+z_1)$. To compute this expression, next, we add 1 and z_1, thus arriving at the second intermediate result

$$z_2 = 1 + z_1.$$

In terms of z_2, the desired expression takes the form

$$y = 1/z_2.$$

This expression can be directly computed.

So, $y(x)$ is a composition of three functions:

- the first function transforms x into z_1,
- the second function transforms z_1 into z_2, and
- the third function transforms z_2 into y.

By comparing the expressions for these functions with the general expressions for invariant functions, we can see that the first function is scale-scale invariant, the second function is shift-shift invariant, and the third function is scale-scale-invariant.

Thus, the desired function is indeed the composition of invariant functions.

Problem 5. Describe a function $\sqrt[3]{x^3 + x^6}$ as a scale-invariant combination of two scale-scale-invariant functions.

Hint: use the fact that $x^6 = (x^2)^3$.

Solution. A scale-invariant combination of two functions $y_1(x)$ and $y_2(x)$ has the form $(y_1^p + y_2^p)^{1/p}$. The above function has $1/p = 1/3$, since $a^{1/3} = \sqrt[3]{a}$. Thus, we have $p = 3$.

So, the above function has the form

$$\sqrt[3]{x^3 + x^6} = (x^3 + x^6)^{1/3}.$$

If we take into account that $x^6 = (x^3)^2$, we conclude that it has the form:

$$\sqrt[3]{x^3 + x^6} = (x^3 + (x^2)^3)^{1/3}.$$

This is the desired form for $y_1(x) = x$ and $y_2(x) = x^2$.

Both functions $y_1(x)$ and $y_2(x) = x^2$ are scale-scale-invariant, so we get the desired representation.

Problem 6. Prove that the family of all linear functions

$$c_0 + c_1 \cdot x$$

is invariant with respect to shifts $x \mapsto x + x_0$ and scalings $x \mapsto \lambda \cdot x$, i.e., that if we substitute $x' = x + x_0$ or $x' = \lambda \cdot x$,into this expression, we still get a linear function – with different coefficients c'_i.

Solution. Let us first show shift-invariance. Here,

$$c_0 + c_1 \cdot x' = c_0 + c_1 \cdot (x + x_0) = c_0 + c_1 \cdot x + c_1 \cdot x_0,$$

i.e., the form

$$c'_0 + c'_1 \cdot x,$$

where we denoted $c'_0 = c_0 + c_1 \cdot x_0$ and $c'_1 = c_1$.

Let us now show scale-invariance. Here,

$$c_0 + c_1 \cdot x' = c_0 + c_1 \cdot (\lambda \cdot x) = c_0 + c_1 \cdot \lambda \cdot x = c_0 + (c_1 \cdot \lambda) \cdot x,$$

i.e., we get

$$c'_0 + c'_1 \cdot x,$$

where $c'_0 = c_0$ and $c'_1 = c_1 \cdot \lambda$.

Problem 7. As alternatives, let us consider families of the type $\{C \cdot f(x)\}_C$, where $f(x)$ is fixed and C can take any value. Let us define shift T_{x_0} as an operation that transforms a family $\{C \cdot f(x)\}_C$ into a new family $\{C \cdot f(x+x_0)\}_C$. Prove that if an optimality criterion on the set of all such alternatives is final and shift-invariant, then each function which from the optimal family has the form $f(x) = A \cdot \exp(a \cdot x)$.

Hint: first, follow the general proof that we had in class about the function optimal with respect to a T-invariant criterion, and then use the result that we proved in class – that every shift-scale-invariant function is described by the above form.

Solution. Let $(<, \sim)$ be a final shift-invariant optimality criterion on the set of all alternatives, and let $F_{\rm opt} = \{C \cdot f_{\rm opt}(x)\}_C$ be the optimal family.

By definition of optimality, this means that $F_{\rm opt}$ is better or of the same quality than all other families, i.e., that for every family F, we have:

$$\text{either } F_{\rm opt} > F \text{ or } F_{\rm opt} \sim F.$$

In particular, this means that for every F and for every x_0, we have

$$F_{\rm opt} > T_{-x_0}(F) \text{ or } F_{\rm opt} \sim T_{-x_0}(F).$$

Due to shift-invariance, we have

$$T_{x_0}(F_{\rm opt}) > T_{x_0}(T_{-x_0}(F)) \text{ or } T_{x_0}(F_{\rm opt}) \sim T_{x_0}(T_{-x_0}(F)).$$

But here, $T_{x_0}(T_{-x_0}(F)) = F$. Thus, for every F, we have

$$T_{x_0}(F_{\rm opt}) > F \text{ or } T_{x_0}(F_{\rm opt}) \sim F.$$

By definition of an optimal alternative, this means that the family $T_{x_0}(F_{\rm opt})$ is optimal. The optimality criterion is final, which means that there is only one optimal family. Thus, $T_{x_0}(F_{\rm opt}) = F_{\rm opt}$.

This means that every function from the family

$$T_{x_0}(F_{\rm opt}) = \{C \cdot f_{\rm opt}(x + x_0)\}_C$$

also belongs to the family $F_{\rm opt}$, i.e., has the form $C \cdot f_{\rm opt}(x)$ for some C. In particular, this is true for the function $f_{\rm opt}(x + x_0)$ from the family $T_{x_0}(F_{\rm opt})$. Thus, for every x_0, there exists a value C depending on x_0 for which

$$f_{\rm opt}(x + x_0) = C(x_0) \cdot f_{\rm opt}(x).$$

We have proven that any function with this property has the form

$$f(x) = A \cdot \exp(a \cdot x).$$

The statement is proven.

Problem 8. Describe your progress on the class project.

Problem 9. Briefly explain what is the purpose of this class.

Solution. The main purpose of computers is to process real-life data, so that we will able to understand the current state of the system and to predict its future behavior.

In some situations – e.g., in basic mechanics – we have fundamental from-first-principles laws that enable us to make the corresponding predictions. However, in many other situations, especially in engineering, we only have approximate empirical formulas. For example, it is not possible to predict, based on the first principles, how pavement will deteriorate with time, or how people will change their opinions about goods.

In such situations, we face the following problems:

- *Why these formulas?* Users are usually reluctant to use purely empirical formulas. Reason: there is no guarantee that theses formulas will work in a new situations. It is therefore desirable to come up with theoretical explanations for these formulas.

- *Maybe these formulas are not the best.* Within these theoretical explanations, are the current formulas most adequate? And if they are not the best, what are the better formulas?

- *What next?* Empirical formulas are usually approximate. If we want a more accurate description, we need more complex, more detailed formulas. Of course, the ultimate test is comparing with the observations and measurement results. In view of the theoretical explanations, what are good candidates for such more complex formulas?

Similarly, in many engineering applications, there are semi-empirical methods for solving the corresponding problems. In such cases, similar problems appear:

- *Why these methods?* Users are usually reluctant to use purely empirical methods, since there is no guarantee that these methods will work in a new situations. It is therefore desirable to come up with theoretical explanations for these methods.

- *Maybe these methods are not the best.* Within these theoretical explanations, are the current methods the most adequate? And if not, what are the better methods?

- *What next?* Empirical methods are usually imperfect. If we want better results, we need more complex, more detailed methods. Of course, the ultimate test is testing these methods on the real data. In view of the theoretical explanations, what are good candidates for such more complex methods?

A natural way to make predictions is to look what happened in similar situations in the past. And what does "similar" mean? It means that some important features of the current situations are the same (or at least almost the same) as the same important features of the past situation. In other words, there may have been some changes between the two situations, but the important features did not change.

In mathematical terms:

- changes are called *transformations*,
- if a feature does not change under a transformation, we say that this feature is *invariant* under this transformation, and
- transformations under which some features are invariant are known as *symmetries*.

Not surprisingly, symmetries and invariances are, at present, one of most effective tools in theoretical physics (and in other disciplines). In line with this reasoning, in this class, we will study the invariance-based approach to solving the above problems.

Here is a simple example. How do we know that: if we drop a pen, then it will start falling down with the acceleration of 9.81 m/sec^2? Because we – and others: repeated this experiment in many different locations, at many moments of time, and always observed the same result.

In this case, the current situation can be obtained from the previous one by shifts in space and time, maybe by rotation. So the observed phenomenon is invariant with respect to all these transformations.

Some transformations studied in physics are very complicated. Example: changing particles to corresponding anti-particles. In this class, we will focus on the simplest and most natural transformations. These transformation come from the fact that: while we want to process the *actual* values of different quantities like acceleration, in practice, we deal with *numerical* values.

For most physical quantities, numerical values depend on the choice of the measuring unit. For example: if instead of meters, we use centimeters, then the same acceleration of 9.81 m/sec^2 becomes described by a different numerical value 981 cm/sec^2.

For some quantities like temperature or time, there is no fixed starting point. So numerical values also depend on what starting point we choose. For example, the difference between Fahrenheit and Celsius scales for measuring temperature is that these two scales use different measuring units and different starting points.

In many situations. there is no fixed measuring unit, the choice of a measuring unit is simply a matter of convention. In this case, it is reasonable to assume that the desired empirical formula has the same form no matter what measuring unit we select.

Of course, each formula relates the values of several quantities; so, if we change the measuring unit for one of the quantities, we need to appropriately change the measuring unit for other quantities. As an example, let us consider the formula $d = v \cdot t$ that describes the traveled distance d as a function of velocity v and time t. This formula remains true whether we measure distance in kilometers or in miles. However, for this formula to remain true we switch from miles to kilometers, we also need to change the units for measuring velocity from miles per hour to kilometers per hour.

Similarly, in some formulas, if we change the starting point for measuring one of the quantities, we may need to appropriately re-scale other quantities.