

Teaching Mathematics for Economists

Eric Hillebrand

Abstract In this chapter, I discuss teaching mathematical tools specifically tailored for economics students. A typical one-semester course in this area seeks to blend a range of topics: from foundational elements of subjects such as linear algebra and multivariate calculus to intermediate areas like convex and real analysis, and further into advanced topics such as dynamic optimization in both continuous and discrete time. This breadth of coverage corresponds to material usually spread across multiple years in traditional mathematics programs. Given the comprehensive nature of these courses, careful selection of topics is essential, balancing numerous trade-offs. I discuss potential course sequences and instructional design choices. I then focus on conceptualizing and explaining mathematical modeling in economics. I reflect on three years of teaching an advanced undergraduate course in mathematical methods online. The latter part of the chapter offers examples and visualizations I've found particularly beneficial for imparting intuition to economics students. They cover a range of topics at different degrees of difficulty and are meant as a resource for instructors in Mathematics for Economists. Among these, I employ the Ramsey model as a recurring example, especially relevant when shaping a mathematical tools course with an orientation towards preparing students for macroeconomic analysis.

1 The task, an ideal course sequence, and the ground truth

From the point of view of a mathematician, teaching mathematics to economists is a near impossible task. At the same time, it is also very much a fun subject to teach in economics programs. Students who have decided to study economics, and not mathematics, should, ideally, have skills in univariate calculus, linear algebra, and basic concepts of multivariate calculus (chain rule, implicit function theorem) comparable to students of mathematics or physics. They should have a solid grasp of

Aarhus University, Department of Economics and Business Economics, Fuglesangsallé 4, 8210 Aarhus, Denmark, e-mail: ehillebrand@econ.au.dk

ordinary differential equations, at least from a practical, solution-oriented perspective. When this is accomplished, elements of convex analysis are necessary, mainly but not exclusively for microeconomic theory, and elements of deterministic dynamic optimization and control in continuous and discrete time for macroeconomic theory. The need-to-haves are, therefore, a curious blend of elements that are at the same time very basic and quite advanced. The list of nice-to-haves is long: real analysis and probability, mathematical statistics (for econometrics), partial differential equations (for finance, mainly), stochastic optimal control (finance, macroeconomics), elements of topology, causal calculus (microeconomics), numerical optimization (all fields), and I am sure that many economists will find it easy to extend this list. In addition, there are a number of areas in economics where the literature is rigorous enough such that graduate students setting out study in these areas should have been introduced to rigorous proof work.

A young mathematician tasked with teaching economics students, remembering their own training where linear algebra was covered in two full semester courses, analysis in three, and ordinary differential equations in one, may therefore think that one semester of linear algebra, one semester of both univariate and multivariate calculus (ideally after the linear algebra course, so that this language can be used), and one with intermediate topics such as differential equations and elements of dynamic optimization and convex analysis may be a good starting point at the undergraduate level. At the graduate level, one course covering analysis in a rigorous manner, for example based on Rudin's book (Rudin, 1976), and some advanced subjects in linear algebra, for example based on Strang's book (Strang, 2023), would provide a basis and a common denominator for the students coming from many different places, and then, in the second year, a choice between two different courses, one in dynamic analysis (differential equations and dynamic optimization) and one in convex analysis and fixed-point theory could cater to the more macro- and micro-oriented students, respectively.¹

In the morning, our mathematician runs into the department head on the way from the parking lot and mentions his rough outline for a course sequence. With a puzzled look, the department head says, "That sounds intriguing. We thought you would be the perfect fit to teach the Chiang course." Before our mathematician has any chance of thinking of a smooth way to rephrase their thought "What on earth is a Chiang course," the department head continues, "Well, we do have two math courses on the books, because we're quantitatively quite strong, one at the undergrad and one at the grad level! If you'd like to revamp a course or design a new one, of course you're welcome to write a proposal." Later that day, the senior econometrician, a very kind and smart man with a knack for mentoring, to whom our young hero understood right away it is important to listen to, drops by. "Hey, I heard you pitched some math courses to the chairman. It's a great thought to do more math, and I would love to see it, but it's going to be an uphill battle to get more courses in than we have right now. And leave the matrix algebra for the graduate level, our undergraduates cannot

¹ Convex analysis and fixed-point theory are also employed to show existence and uniqueness of value functions in dynamic optimization, and there are mathematics-for-economists textbooks that cover this, for example, Corbae et al. (2009).

stomach it, believe me, I tried. Same goes for complex numbers, and we don't really need them in economics, anyway."

Thus goes up in smoke a dream of our mathematician to share with their students their own undergraduate epiphany that once you allow for higher dimensions, for example 2×2 , the need for i disappears, since the equation

$$X^2 = -I$$

has the perfectly integer solution

$$X = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{Z}^{2 \times 2}.$$

Lo and behold, the eigenvalues of this matrix are i and $-i$, but you don't need to insist on computing them.

The view that linear algebra cannot be communicated to undergraduate economics students is very common. It pervades the design of many textbooks, also otherwise excellent ones, such as Sydsæter et al. (2022) and Sydsæter et al. (2008).² I think that this view is one of the major hindrances for a better math education within economics. It drastically reduces the speed with which progress can be made towards the material that is necessary for understanding and contributing to the economics literature. Or that would enable the economics student to appreciate when neighboring disciplines, such as engineering, talk about exactly the same things as economists do (Luenberger, 1969). The situation results in many serious students, who are ambitious about an academic career in economics, resorting to taking courses at the mathematics departments. On the other hand, maybe it is also better that way. There is a certain mathematics field culture that is very useful to have experienced as an economist working with rigorous material, and it is probably best when it comes from the horse's mouth. The serious student quickly realizes that even the supposedly basic material in linear algebra, analysis, and differential equations presents itself at a completely different depth and brilliance when studied carefully. If they made the commendable decision to take some of the undergraduate courses at the mathematics department, for example as a first-year economics Ph.D. student, they are then often met with the parodic verdict that they cannot get graduate credit for taking undergraduate classes. . .

Returning to linear algebra, should it really be beyond the capacities of abstraction of a 19- to 23-year old student of economics that a vector is a point in a properly defined space that behaves under addition with other vectors and multiplication with scalars as you would expect it? That multiplication of a vector with a matrix, for example the product $y = f(x) = Ax$, where

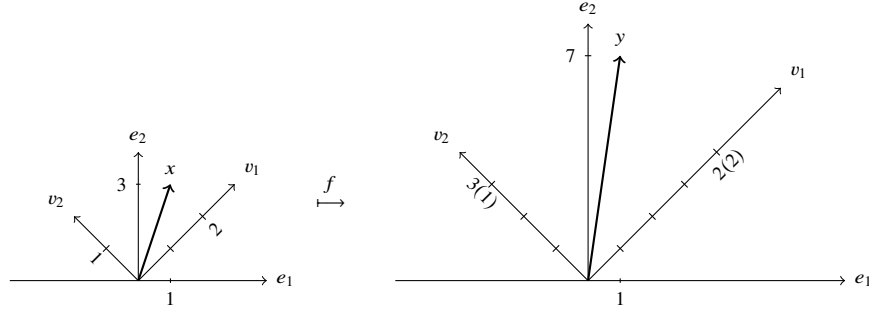
$$A = \frac{1}{2} \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

is given by the two rows of addition and multiplication operations

² A notable exception is Simon and Blume (1994).

$$\begin{bmatrix} \frac{5}{2} \cdot 1 - \frac{1}{2} \cdot 3 \\ -\frac{1}{2} \cdot 1 + \frac{5}{2} \cdot 3 \end{bmatrix}?$$

That every multiplication of a vector x with a matrix A amounts to a rotation and a scaling of x ?



That angle of rotation and scaling factor are, however, in general not independent of x , that they can vary from vector to vector?³ That despite this dependence of the scaling and rotation on the argument vector of the function, many matrices allow for a transformation of the coordinate system such that the scaling in that transformed coordinate system *along those axes* is always the same? In our example, the transformed coordinate system is given by the vectors

$$\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right).$$

In the multiplication $y = Ax$, first the old coordinates $(1, 3)$ of x are switched to the new coordinates $(2, 1)$ with respect to this system. Then, the new coordinates are stretched by the eigenvalues, the first coordinate by the first eigenvalue, the second coordinate by the second eigenvalue, to $(4, 3)$. These are the new coordinates of y . Finally, the new coordinates are switched to old coordinates $(1, 7)$.

³ That only specific matrices always have the same rotational angle and the same scaling independent of x ? $A = \begin{bmatrix} \alpha \cos \theta & -\alpha \sin \theta \\ \alpha \sin \theta & \alpha \cos \theta \end{bmatrix}$, $\alpha \in \mathbb{R}$ scaling factor, $\theta \in [0, 2\pi]$ rotational angle.

$$\begin{aligned}
y = Ax &= \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_P \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}}_\Lambda \underbrace{\frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{P^{-1}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_P \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}}_\Lambda \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\
&= \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_P \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}
\end{aligned}$$

I'm convinced that this is within the geometric and algebraic reach of an undergraduate economics student. They may complain about the algorithmic way to arrive at the eigenvalue decomposition, but it involves finding the roots of a polynomial and solving a system of equations, all things they have seen before in some way or another.

Be that as it may, our young mathematician now has to make do with the one-semester undergrad and the one-semester graduate course plus the senior econometrician's advice. At the undergrad level, they will end up teaching some version of multivariate calculus with two or at most three variables, writing scalar partial derivatives, not gradients or Jacobians. They will likely do some version of optimization under constraints but cannot really talk about constraint qualification. Some review of integration of functions of one variable will likely be done. Maybe some scalar first-order ordinary differential equations are covered, certainly not systems or higher-order, where there wait so many deeply satisfying insights into the relations of ODEs with linear algebra. If the young mathematician is very organized and does not follow their impulse to delve too deeply into the subjects, some dynamic optimization can be covered, with some luck, before the semester runs out. Either in continuous or in discrete time, most likely not both, and of course scalar only, likely only deterministic and not stochastic. The students will still find the course dense and challenging. The chalkboard is full of math, after all.

Most mathematics-for-economists textbooks represent some compromise along these lines. Some opt for the voluminous or even two-volume approach (Simon and Blume, 1994; Hoy et al., 2001; Klein, 2002; Sydsæter et al., 2022), others for the small-and-lean approach (Weintraub, 1982; Novshek, 1993). In the graduate course, remedial linear algebra now needs to be taught. This takes easily a third of the semester. There's still time to teach some real analysis, proof techniques, advanced topics in convex analysis, and in dynamic optimization. Similarly to the undergraduate situation, a balance has to be struck, and the specific solution usually ends up depending on a combination of the popular specializations that Ph.D. students take in the program at hand and the personal preferences of the instructor. Comparing graduate texts, it quickly becomes clear that the subject can be treated "right" in many different ways, see, for example Ok (2007), tailored towards econometrics, Corbae

et al. (2009), tailored towards macroeconomics, and Sydsæter et al. (2008), trying to balance macro- and microeconomics, leaning towards dynamic optimization. Both at undergrad and grad level, single one-semester courses prevail, which have some sort of umbrella title such as “Mathematics for Economics”, “Mathematical methods in economic analysis” or similar, and do not branch out into subjects or purposes. Together with the common perception that textbooks should be relatively new, even though the subject material taught has essentially not changed in over 50 years, this makes it difficult to assign a number of wonderful monographs with a bit more specific titles (for example, Seierstad and Sydsæter, 1987; Dixit, 1990; Intriligator, 2002; Dhrymes, 2013, and this is certainly not an exhaustive list).⁴

2 Mathematical modeling in economics

This essay starts from the assumption that students of economics should be taught a level of mathematics that enables them to comprehend (undergraduate and master graduate) and contribute to (Ph.D. graduate) parts of the academic economics literature. The ideal course sequence proposed above operates from this basic assumption.⁵ Whether the mathematical sophistication employed in the economics literature is always useful for good economics is a different question that reasonable people can argue about, and I do not address that question here. Inevitably, at some point, the math-for-econ instructor is confronted with the student question what all these mathematical models are good for, if the instructor “believes” in them, if they are “true”, or some variant of this. This is a good question that deserves encouragement.

The probably most famous answer to this question is George E.P. Box’ (Box, 1979):

“All models are wrong but some are useful,”

with an entire Wikipedia web site devoted to it.⁶

The quote pointedly describes the trade-off between accuracy and usefulness of a model. In my own lectures, I like to illustrate this trade-off with two extreme examples: On the one extreme, imagine mythological explanations for why to seek higher ground when the water at the beach recedes.⁷ The mental model used here

⁴ M.D. Intriligator noted in the 2002 SIAM Classics edition of his 1971 “Mathematical Optimization and Economic Theory” that the basic methods covered remain the same as in 1971, and that the only change he would have made was putting more emphasis on duality, reflecting developments in mathematical economics.

⁵ The subject and the courses discussed here are “mathematics for economists/economics”, i.e., tools courses, and not “mathematical economics.” There is no universally agreed upon distinction, and a good deal of confusion, but it makes sense to me to reserve the latter term, as many economists do, for formal economic equilibrium theory, represented by texts such as Debreu (1959), Arrow (1963), Arrow and Intriligator (1981, 1982, 1986), or, for example, Stigum (1990).

⁶ https://en.wikipedia.org/wiki/All_models_are_wrong

⁷ http://news.bbc.co.uk/2/hi/south_asia/4181855.stm

is almost certainly entirely inaccurate, but tremendously useful. I'd certainly prefer surviving with it over drowning with differential equations, because I, again, forgot the integration constant. On the other extreme, consider a perfectly accurate model, such as a map of a city at scale 1:1 and in three dimensions. One is as lost in the map as one is in the actual city, so it is entirely useless.⁸ There are other dimensions that play into the trade-off, such as the state of understanding of the causal processes at work, the effort required to run the model, and the purpose of the model.

John von Neumann defined science as model building only (von Neumann, 1995):

"The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work – that is, correctly to describe phenomena from a reasonably wide area."

This seems overly restrictive to me, but it focuses on interesting aspects of model building: Explanations and interpretations of causal chains are usually not the result of models, they are premises of the model building cycle. The model is a metaphor of the causal processes at work according to current understanding, and it helps to organize thinking and to put features of interest into sharp relief. This aids communication of ideas between researchers (and between teachers and students), and it helps to pose sharp questions. If the model can predict, great, but this is often not the first interest in the model. It is more important that the model can describe the phenomenon of interest with a minimum of assumptions. In economics, this minimum is often referred to as "from first principles." Examples of first principles are diminishing marginal returns, concave or quasi-concave utility functions, non-zero time preferences, and so on. Modelers find it satisfying if an empirical regularity can be recreated in the model with the minimal set of assumptions. Contrary to von Neumann's dictum, they then like to say that the model "explains" the regularity.

In statistical and econometric models, the empirical reality to be captured are the data. In structural models, it is again the minimum set of assumptions borrowed from economic theory and phrased as a statistical model by adding error terms subject to simple distribution assumptions that is seen as desirable. The litmus test is whether the residuals of the estimated model satisfy the distributional assumptions according to diagnostic statistics. In reduced-form models, it is minimal complexity, or parsimony, that is strived for. Parsimony is always in some sense a minimum number of parameters. They can be associated with explanatory variables, lags of variables, deterministic or stochastic features of the model, etc. Again, residual diagnostics consistent with the model's distributional assumptions are aimed for

⁸ I think that we are currently observing a perfectly accurate model of money in the emergence of cryptocurrencies. These serve all three functions of money and therefore satisfy the standard textbook definition accurately. However, another definition of money that one often hears when talking with bankers is that money is a liability of a money-issuing, usually central bank, i.e., that there is a creditworthy institution that has the money issue on the passive side of the balance sheet. If the failure of cryptocurrencies to satisfy this definition becomes problematic, this will at the same time demonstrate the uselessness of cryptocurrencies as a model for money, because they do not improve our understanding of money.

with the minimum number of parameters. Never is the model conceptualized as an accurate description of reality, not even when the “nonlinearities” are taken into account. To take the model literally is to not take it seriously (Thompson, 2022). The choice which features of reality to emphasize in a model and which ones to tone down depends on a mixture of all the aspects mentioned above, plus, to no small extent, the skill and experience of the modeler. Often, statements to the effect that modeling is an art rather than a science are made. Modeling is certainly a creative process.

As a caveat, my vague understanding is that philosophy of science has come a long way since the people who I’m citing here were active and helped shape the epistemological background of my generation (Schindler, 2018). On a lighter note, modeling is taken seriously in economics to the point of inviting self-parody. I recommend the unparalleled “Life Among the Econ” (Leijonhufvud, 1973):

“The Econ word for caste is ‘field.’ Caste is extremely important to the self-image and sense of identity of the Econ, and the adult male meeting a stranger will always introduce himself with the phrase ‘Such-and-such is my field.’ [...] A comparison of status relationships in the different ‘fields’ shows a definite common pattern. The dominant feature, which makes status relations among the Econ of unique interest to the serious student, is the way that status is tied to the manufacture of certain types of implements, called ‘modls.’ The status of the adult male is determined by his skill at making the ‘modl’ of his ‘field.’ The facts (a) that the Econ are highly status-motivated, (b) that status is only to be achieved by making ‘modls,’ and (c) that most of these ‘modls’ seem to be of little or no practical use probably accounts for the backwardness and abject cultural poverty of the tribe.”

3 Teaching Mathematics for Economists online

At Aarhus University, at the time of writing, we are in the relatively luxurious position to have two-and-a-half courses in mathematics for economists, one-and-a-half at the undergraduate level and one at the graduate level. The half course (5 ECTS) is the first half of “Principles of Mathematics and Statistics,” taught in Danish in the first semester, and the mathematics part is based on selected chapters from the book by Sydsæter et al. (2022). There are about 200 students. The full course (10 ECTS) is called “Mathematics for Economists,” which I had renamed in a long process from “Mathematical Economics” after I took over in 2014. It is taught in Danish in the third undergraduate semester, and it is based on selected chapters from the book by Sydsæter et al. (2008). There are about 160 students. In this course, I cover linear algebra in three weeks, multivariate calculus in the next three weeks, static optimization (global, local, and constrained extrema) in one-and-a-half weeks, integration of functions of several variables (one week), first-order ordinary differential equations (one week), second order linear differential equations and two-dimensional systems (one week), control theory (two weeks), and difference equations and dynamic optimization in discrete time (one-and-a-half weeks), for a total of fourteen weeks. The exam format is a 20-minute oral exam. We screen the

incoming undergraduate students for mathematics; the Danish high-school system allows for taking math at different levels, and we require the highest.

The Ph.D. course was taught for years, in English, by a colleague from the mathematics department, who used Rudin (1976). There are about 10 students. After his retirement, the course has been taught by different colleagues who have worked with different texts (Boyd and Vandenberghe, 2004; Johnsonbaugh and Pfaffenberger, 2010; Ok, 2007; Resnick, 2014). The emphasis on real analysis represents a least common denominator for all fields, since it directs the focus towards rigorous proof training. The probability text was used supplementarily. We have a number of international Ph.D. students, but the majority are Danish with their undergraduate degree from Aarhus, so they had the third-semester undergraduate course.

This was also the situation in 2020, when COVID-19 hit. I had been toying with the idea to take the third-semester course with about 160 students online for a long time before that, however. As noted above, the subject matter has not changed substantially in decades, and it certainly does not change from year to year. This makes the course a near perfect candidate for recording videos and experimenting with an online format. I had some experience with recording videos on mathematics for economists from my YouTube channel, which I had started sometime in 2016.⁹ I took the situation in 2020 as an opportunity to give it a try. The easiest part was to choose a video format. I went for a three-part layout, with half of the screen filled with a projection of my tablet to write on, a quarter of the screen filled with the slides and another quarter with a headshot video of me. This way, I figured, the students would have the main lecture experience in a nutshell: the instructor, the blackboard, the slides.¹⁰

In 2020, the lockdown situation necessitated also taking the tutorials, run by student teaching assistants, online. This was the hairier problem to solve. In the physical format, I had been giving weekly homework sets with exercises, some from the book, some of my own, and in the in-person tutorials, the students had been discussing these exercises. For the online format, I decided to drastically reduce the number of mathematical exercises to work through and instead added an exercise whereby each working group of three or four students had to prepare an eight-minute video with a presentation of the main elements of the week's material. The students recorded these videos with their smartphones, for example by laying them on a book stack with the camera aimed at a sheet of paper. The eight minutes were chosen because this is about the time a student presents in the 20-minute exam. The rest of the time is spent with questions and answers.

In the online tutorial, half of the time was spent watching and discussing the videos and the other half discussing the mathematical exercises. My main intention was to prepare the students for the likely scenario of online oral exams, which then also came to pass, both in 2020 and in 2021, due to government-ordered lockdowns. (The exams are in January.) In preparing the videos, the students could establish a routine with their improvised recording setups and with the situation of explaining

⁹ https://www.youtube.com/channel/UCJjiXpAzy_L6uneX5uoTihQ

¹⁰ An example, in Danish, can be seen here: <https://www.youtube.com/watch?v=HDXOZnMLaes>.

math to a camera and a microphone.¹¹ Another intention was to give the students the opportunity to see other students in exam-like presentations, and to make use of the common asymmetry that one is better at judging than at performing. The hope was that they would catch some of the common mistakes I had seen in the exams early on, when other students made them in the tutorials. This worked to a degree, but meaningful feedback on the videos between the different working groups remained an exception. The students clearly did not feel comfortable critically discussing their peer's videos and when asked resorted to short and generally positive and encouraging statements ("great video!"). I got feedback myself from our Center for Educational Development that in their experience, getting students to give each other meaningful feedback takes doing workshops with them.

I ended up running the online format for three consecutive years, from 2020 to 2022. In 2021 and 2022, the tutorials were in person. As noted above, the oral exams had to be held online in 2020 and 2021. I met with the teaching assistants once a week to discuss the ongoing at the tutorials and the exercises. In the in-person format, there had been years where, to my embarrassment, I had difficulties remembering the TA's names by the end of the semester. In the online format, this became unthinkable. You could have woken me up in the middle of the night and I could have given you the full names of my (usually four) TAs. This was due in no small part to the realization that by choosing the online format, I had transferred the main teacher role in the teacher-student relationship to them. They were the ones who met with the students once a week and who were the course's "face" to the students. They saw the students work, answered their questions, and only related questions and issues they could not solve themselves to me. Their online office hours were heavily frequented, whereas my own online office hours were largely perfunctory.

To keep some traction with the students and to stay in touch, I visited the tutorials on a sampling basis, both in the online and in-person tutorial versions. Since there usually are eight tutorial groups of about 20 students, I could visit each at most twice per semester. I got very enthusiastic about this process. In the in-person format, I lectured in front of 160 students in a large hall, and interaction is limited to the two handful of students who come up in the break or after the lecture. I only see the students work, and learn their names, in the oral exam at the end of the semester. In the online format, I saw them working on problems on the blackboard from the first week on, identified the good students throughout the semester (TA material for later), and generally had the feeling of being much closer to the students. It took me a while to understand that this feeling was thoroughly asymmetric. In the in-person format, the students see me twice a week for 90 minutes, hear me talk, watch me work, and get myself chalk-dusty. Now they saw me in a small video window on YouTube and otherwise at best twice in the tutorial, mostly sitting and listening and then giving feedback on presentations, likely not their own, and making some general comments, and that was it. They felt I was far away.

¹¹ The level of student's creativity and ingenuity was something to behold. A stellar moment was a student who sat in her kitchen for the online oral exam with the smartphone camera aimed at her white refrigerator door, on which she answered our questions with a whiteboard marker impeccably. It was beautiful.

In the oral exams, the online students fared substantially worse. They were maybe marginally better at presenting what they had prepared, but they were generally much worse at answering anything that lay outside of their preparation. This was also often true for obvious and predictable questions, such as when a subject was called “Second-order ordinary differential equations and systems in the plane,” and they had prepared a presentation on second-order equations and were then surprised when I asked into systems. This also happened in spite of explicit and coordinated warnings to this effect in the tutorials. I was puzzled about the possible causes, and I still haven’t come to a satisfying conclusion. My working hypothesis is that the regular presentations in the tutorials, which were a much less tense situation compared to the exam, only followed up by friendly feedback from the student TA, had framed the exam as just another one of those. But I’m not sure.

Another effect I observed has also been documented in a large-scale randomized controlled trial in the literature (Cacault et al., 2021). The good students thrived in the online format. They were intrinsically motivated and interested in engaging the material, and so they could enjoy the flexibility the online format gave them. They liked it and were as good as ever. The not so good students, however, had a worse outcome than in the in-person format. They struggled to keep up with the weekly flow of material, and in informal discussions with the student TAs, a higher-than-normal percentage of students gave the feedback that they were lost. They did not show up for my online office hours, probably for fear of embarrassment, and the student TA office hours were a reaction to this phenomenon. Still, only a few dared come to their TAs, and even though the student TA office hours were a success, they were also much frequented by the good students who came to discuss finer points, for which there also should be time, of course. Again, I can only hypothesize as to the possible causes. One is that upon hitting a rough patch in working with material, it is, of course, much easier to close down a laptop than it is to leave a room full of people. But also in the in-person format most of the deeper understanding is happening at home, so this is only a partial explanation. Another element may be that in the in-person format, a struggling student can physically see peers successfully engaging in the material, and this may give them a much needed kick, maybe out of a sense of competitiveness, maybe because they think, if they can do it, I should be able, too, maybe simply because it sets a certain tone and positive attitude. Wherever the effect was originating, it was clearly discernible.

In my own time economy, after the first year, the online course took almost, but not quite the same space as the in-person version. In the first year, recording the videos took an inordinate amount of time, and for about two months, I was not doing anything else. In the following years, visiting the tutorials, holding online office hours, engaging with the TAs at a deeper level, and fielding student emails all amounted to just slightly less work than the in-person version. The total time commitment with students and TAs was about the same, but I did not have to prepare classes.

In the fall of 2023, I switched back to the in-person format. The reason was a combination of all of the above, but the main motivation was the fact that I became a professor to teach and to research. I enjoy teaching. In the online format, my role

was reduced to that of a teaching organizer and facilitator. I did many things that had to do with teaching, but as noted above, the actual teachers were the TAs. Not having to prepare classes, I realized to my chagrin that I became a bit rusty in the subject material. Returning to the in-person format, I eliminated the presentation exercise from the homework sets and increased the number of problems to solve pencil to paper, or stylus to tablet. The students became better again at answering exam questions outside of their preparations, at least they were not surprised. There is a time trend, however, that is pointing in the wrong direction. Occasionally, I get the comment that written exercises are not the best preparation for oral exams, because that seems a true thing to say. This is difficult to reply to without telling the long story, but the answer is, as I now know, that they actually are. Some students ask if I could make the old videos available, which I do not think is a good idea. The one linked above is made accessible solely for the purpose of this paper.

4 Examples and exercises

In this section, I present selected examples and exercises. They are meant as a resource for instructors in Mathematics for Economists and may be freely used. My github repository houses the Latex code for this section, including the tikz bits for the illustrations.

4.1 The geometry of the determinant

“The determinant of a square matrix A , denoted by $|A|$, is a uniquely defined scalar (number) associated with that matrix.” (Chiang, 1984, p. 93)

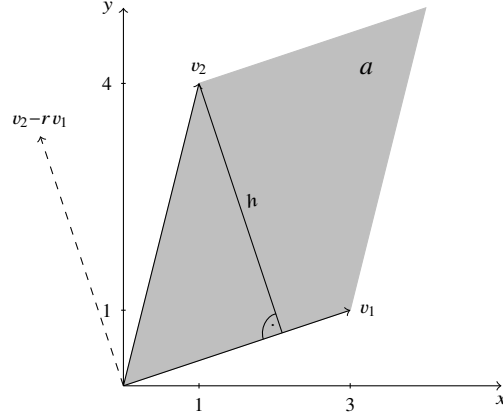
The determinant of an $n \times n$ square matrix is the n -dimensional volume of the parallelepiped spanned by the column (or row) vectors of that matrix. There is a sign convention for this volume similar to sign conventions in integration. Thus, in the case of $n = 2$, the determinant is the area of a parallelogram.

4.1.1 The determinant of a 2×2 matrix

In this example, I show how the determinant of a 2×2 matrix describes the area of the parallelogram spanned by the column vectors using arguments from geometry and trigonometric functions that the students should have seen in high school. I also make use of the inner, or scalar product, which I introduce right before determinants.

Consider the matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix} =: [v_1 \ v_2].$$



The columns v_1 and v_2 of A span a parallelogram $\{r_1 v_1 + r_2 v_2 | 0 \leq r_{1,2} \leq 1\}$. To determine the area a of the parallelogram, we need to find the height $\|h\|$. We have that $\|h\| = \|v_2 - r v_1\|$ for some $r \in \mathbb{R}$, and $v_2 - r v_1$ is perpendicular to v_1 . This results in the following condition for r .

$$\begin{aligned} \langle h, v_1 \rangle &= \langle v_2 - r v_1, v_1 \rangle = \langle v_2, v_1 \rangle - r \|v_1\|^2 \stackrel{!}{=} 0 \\ &= 7 - 10r \implies r = \frac{7}{10}, \end{aligned}$$

where $\langle x, y \rangle = x^T y = x_1 y_1 + x_2 y_2$ denotes the inner product, or scalar product, of the vectors x and y . Then,

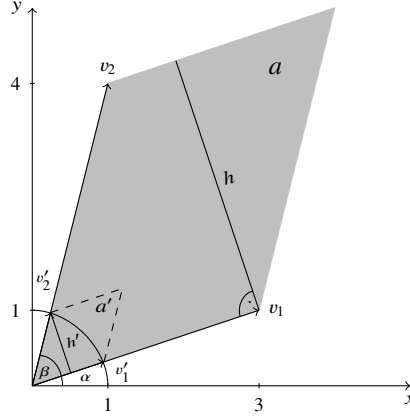
$$\begin{aligned} h &= v_2 - r v_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} - \frac{7}{10} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -1.1 \\ 3.3 \end{bmatrix}, \\ \|h\| &= \|v_2 - r v_1\| = \sqrt{1.1^2 + 3.3^2}, \\ a &= \|v_1\| \|h\| = \sqrt{10} \sqrt{1.1^2 + 3.3^2} = 11, \end{aligned}$$

and we have found the area of the parallelogram to be 11 units.

Alternatively, define

$$v'_1 := \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad v'_2 := \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{17}} \begin{bmatrix} 1 \\ 4 \end{bmatrix}.$$

The vectors v'_1 and v'_2 have unit length. Consider the parallelogram spanned by v'_1 and v'_2 and denote its area a' .



The coordinates of the unit vectors are $v'_1 = (\cos \alpha, \sin \alpha)^T$ and $v'_2 = (\cos \beta, \sin \beta)^T$. Since v_1 is a multiple of v'_1 , and v_2 is a multiple of v'_2 , we can write for $\rho = \sqrt{17}$, $\sigma = \sqrt{10}$,

$$v_1 = \rho v'_1 = \rho \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix} = \|v_1\| v'_1 = \sqrt{17} v'_1,$$

$$v_2 = \sigma v'_2 = \sigma \begin{bmatrix} \cos \beta \\ \sin \beta \end{bmatrix} = \|v_2\| v'_2 = \sqrt{10} v'_2.$$

The length of h' is given by

$$\|h'\| = \frac{\|h'\|}{\|v'_2\|} = \sin(\beta - \alpha).$$

The area a' is therefore

$$\begin{aligned} a' &= \|h'\| \|v'_1\| = \|h'\| = \sin(\beta - \alpha), \\ &= \cos \alpha \sin \beta - \cos \beta \sin \alpha, \end{aligned}$$

Define the matrix

$$A' = \begin{bmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{bmatrix} = \begin{bmatrix} \cos \alpha & \cos \beta \\ \sin \alpha & \sin \beta \end{bmatrix} = [v'_1 \ v'_2],$$

then we notice that

$$a' = \cos \alpha \sin \beta - \cos \beta \sin \alpha = a'_{11} a'_{22} - a'_{21} a'_{12}.$$

Consider the matrix A that spans the larger parallelogram

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \rho \cos \alpha & \sigma \cos \beta \\ \rho \sin \alpha & \sigma \sin \beta \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}.$$

Then we still have that $\sin(\beta - \alpha) = \|h\|/\|v_2\|$ and

$$\begin{aligned} a &= \|v_1\| \|h\| = \|v_1\| \|v_2\| \sin(\beta - \alpha), \\ &= \rho \sigma (\cos \alpha \sin \beta - \cos \beta \sin \alpha), \\ &= \rho \cos \alpha \sigma \sin \beta - \sigma \cos \beta \rho \sin \alpha, \\ &= a_{11}a_{22} - a_{21}a_{12} = 3 \cdot 4 - 1 \cdot 1 = 11. \end{aligned}$$

For the students to whom the sum formula

$$\sin(\beta - \alpha) = \cos \alpha \sin \beta - \cos \beta \sin \alpha$$

is unknown or feels too much pulled out of the hat, I add an exercise/example when we cover complex numbers:

$$\begin{aligned} e^{(\beta - \alpha)i} &= \cos(\beta - \alpha) + i \sin(\beta - \alpha), \\ &= e^{\beta i} e^{-\alpha i} = (\cos \beta + i \sin \beta)(\cos \alpha - i \sin \alpha), \\ &= (\cos \alpha \cos \beta + \sin \alpha \sin \beta) + i(\cos \alpha \sin \beta - \sin \alpha \cos \beta), \end{aligned}$$

and thus

$$\begin{aligned} \cos(\beta - \alpha) &= \operatorname{Re}(e^{(\beta - \alpha)i}) = \cos \alpha \cos \beta + \sin \alpha \sin \beta, \\ \sin(\beta - \alpha) &= \operatorname{Im}(e^{(\beta - \alpha)i}) = \cos \alpha \sin \beta - \sin \alpha \cos \beta. \end{aligned}$$

If you'd like to follow the senior econometrician's advise and avoid complex numbers altogether, there are also beautiful visual proofs of this relation, for example in Nelsen (2000), p. 43. For the students to whom the Euler formula $e^{ix} = \cos x + i \sin x$ is unknown or feels foreign, I add an exercise when we cover Taylor expansions:

1. Let $f(x) = \sin(x)$. Use the Taylor-formula to show that

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}.$$

One finds

$$f(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \dots$$

2. Find a corresponding expression for $\cos(x)$.

$$f(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots$$

3. Use the two results to show that

$$\exp(ix) = \cos(x) + i \sin(x),$$

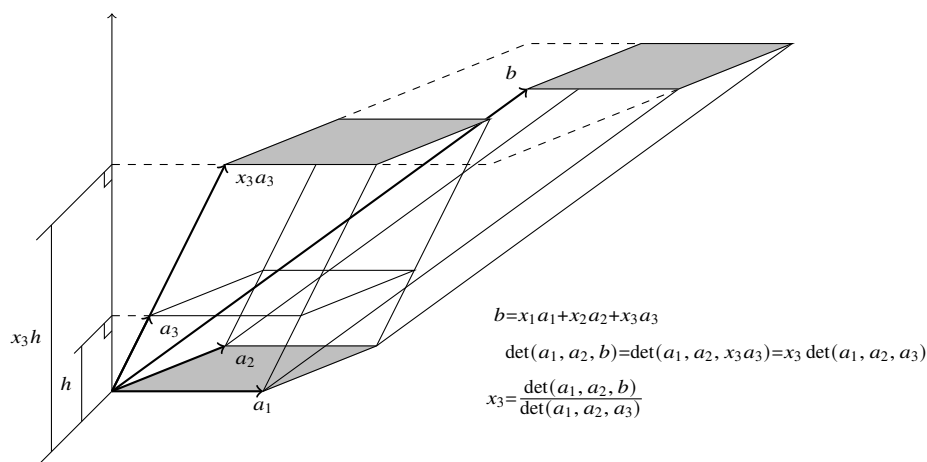
where i is the imaginary unit satisfying $i^2 = -1$.

One finds

$$\begin{aligned} \exp(ix) &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - \dots \\ &= \left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots\right) \\ &\quad + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \dots\right). \end{aligned}$$

4.1.2 Cramer's formula

Equipped with this geometric intuition of the determinant, the students can now appreciate this wonderful visualization, which I found in Nelsen (2000).



From Nelsen (2000), "Proofs without words."

4.1.3 Polynomials as determinants/volumes

This example shows that all polynomials are at the same time determinants of a matrix and thus volumes of a parallelepiped. The students will have to work with the characteristic polynomial in the eigenvalue problem, and this example establishes the close link between matrices, determinants, and polynomials. Consider the matrix

$$A = \begin{bmatrix} a_{n-1} + x & a_{n-2} & a_{n-3} & \cdots & a_1 & a_0 \\ -1 & x & 0 & \cdots & 0 & 0 \\ 0 & -1 & x & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x & 0 \\ 0 & 0 & 0 & \cdots & -1 & x \end{bmatrix} = xI - \begin{bmatrix} -a_{n-1} & -a_{n-2} & -a_{n-3} & \cdots & -a_1 & -a_0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Expand the determinant along the first row

$$\begin{aligned} \det A &= (a_{n-1} + x) \det \begin{bmatrix} x & 0 & \cdots & 0 & 0 \\ -1 & x & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & x & 0 \\ 0 & 0 & \cdots & -1 & x \end{bmatrix} - a_{n-2} \det \begin{bmatrix} -1 & 0 & \cdots & 0 & 0 \\ 0 & x & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & x & 0 \\ 0 & 0 & \cdots & -1 & x \end{bmatrix} \\ &\quad + a_{n-3} \det \begin{bmatrix} -1 & x & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & x & 0 \\ 0 & 0 & \cdots & -1 & x \end{bmatrix} + \dots + (-1)^{n+1} a_0 \det \begin{bmatrix} -1 & x & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & x \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}. \end{aligned}$$

The right-hand side determinants are easily evaluated since they reduce to determinants of upper and lower triangular matrices.

$$\begin{aligned} \det A &= x^n + a_{n-1}x^{n-1} - a_{n-2}(-1)x^{n-2} + \dots + (-1)^{2n}a_0, \\ &= x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0. \end{aligned}$$

Therefore, any n -th degree polynomial is the volume of the parallelepiped spanned by the columns of the matrix A . The GeoGebra object <https://www.geogebra.org/3d/n2ppshze> shows the case $n = 3$ with $a_2 = 1$, $a_1 = 1$, and $a_0 = 1$ and the column vectors for the argument values $x = -1$ and $x = 3$.

4.2 Eigenvalues and Lagrange Multipliers

4.2.1 The gradient of a quadratic symmetric form

This example or exercise shows how to apply the definition of differentiability of a function from \mathbb{R}^n to \mathbb{R}^m to symmetric bilinear forms. I find it particularly intuitive because of the close resemblance of the gradient with the univariate situation. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Consider the quadratic symmetric form

$$q_A(x) = x^T A x = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \langle x, A x \rangle.$$

Then,

$$\begin{aligned}
q_A(x+h) &= \langle x+h, A(x+h) \rangle = \langle x+h, Ax+Ah \rangle \\
&= \langle x, Ax \rangle + \langle x, Ah \rangle + \langle h, Ax \rangle + \langle h, Ah \rangle, \\
\langle x, Ah \rangle &= x^T Ah = (x^T Ah)^T = h^T Ax = \langle h, Ax \rangle.
\end{aligned}$$

Therefore,

$$\begin{aligned}
q_A(x+h) &= \langle x, Ax \rangle + 2\langle h, Ax \rangle + \langle h, Ah \rangle, \\
&= q_A(x) + \langle 2Ax, h \rangle + \langle h, Ah \rangle,
\end{aligned}$$

where

$$\lim_{h \rightarrow 0} \frac{\langle h, Ah \rangle}{\|h\|} = 0.$$

The definition of differentiability of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is that there is a representation

$$f(x+h) = f(x) + Ch + R(h), \quad x, h \in \mathbb{R}^n, C \in \mathbb{R}^{m \times n}, \lim_{h \rightarrow 0} R(h)/\|h\| = 0.$$

The result is that quadratic symmetric forms $q_A : \mathbb{R}^n \rightarrow \mathbb{R}$ are totally differentiable with

$$C^T = \nabla q_A(x) = 2Ax \in \mathbb{R}^{n \times 1}.$$

Loosely speaking, the quadratic form $x^T Ax$ is the multivariate version of the function $f(x) = ax^2$, and so this result is not surprising.

4.2.2 Extrema of a quadratic symmetric form on the unit sphere

This example extends the insight into symmetric bilinear forms from the derivation above to an optimization problem. It is an example where Lagrange multipliers turn out to be eigenvalues. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Find the extrema of $q_A(x) = x^T Ax$ on the unit sphere

$$S = \{x \in \mathbb{R}^n \text{ such that } \|x\|^2 = \langle x, x \rangle = 1\}.$$

Since q_A is a continuous function and S a compact set, the extrema exist. Write the constraint as

$$g(x) = \langle x, x \rangle - 1 = \sum_{i=1}^n x_i^2 - 1 = 0.$$

The partial derivatives and the gradient of the constraint are

$$\frac{\partial g}{\partial x_i} = 2x_i, \quad \nabla g(x) = 2x.$$

According to the Lagrange multiplier theorem, the necessary condition is

$$\begin{aligned}\nabla q_A &= \lambda \nabla g, \\ 2Ax &= \lambda 2x,\end{aligned}$$

or

$$Ax = \lambda x,$$

which we recognize as the eigenvalue problem. Denote the minimizer of q_A on S by x_0 and the maximizer by x_1 . Then,

$$\begin{aligned}q_A(x_0) &= \langle x_0, Ax_0 \rangle = \langle x_0, \lambda_0 x_0 \rangle = \lambda_0 \langle x_0, x_0 \rangle \\ &= \lambda_0 \text{ smallest eigenvalue,} \\ q_A(x_1) &= \langle x_1, Ax_1 \rangle = \langle x_1, \lambda_1 x_1 \rangle = \lambda_1 \langle x_1, x_1 \rangle \\ &= \lambda_1 \text{ largest eigenvalue.}\end{aligned}$$

It follows that the smallest and the largest eigenvalue of A are the minimum and the maximum of $q_A(x) = x^T Ax$ on the unit sphere S and at the same time the Lagrange multipliers of this constrained optimization problem.

4.3 The airborne fraction of anthropogenic CO₂ emissions

This is an example of a first-order linear differential equation with constant homogeneous coefficient and time-dependent inhomogeneous term. A common assumption in climate science is that the terrestrial biosphere (plants) and the ocean absorb atmospheric CO₂ linearly,

$$\begin{aligned}F_{oc}(t) &= \frac{1}{\tau_{oc}} C(t), \\ F_{ld}(t) &= \frac{1}{\tau_{ld}} C(t),\end{aligned}$$

where $F_{oc,ld}$ denotes absorption by ocean and land (F -flux, oc -ocean, ld -land), and the positive numbers τ_{oc} and τ_{ld} denote the average time it takes for land and ocean to absorb one unit carbon C of atmospheric CO₂.

Under the assumption that anthropogenic CO₂ emissions are sufficiently well-described by an exponential function (which is roughly true looking at the total period from industrialization to today, but on the more recent sample, a case can be made that they are linear, see Bennedsen et al. (2023)),

$$f(t) = f_0 e^{dt}, \quad d > 0,$$

it can be shown that the so-called “*airborne fraction*”

$$AF = \frac{1}{f(t)} \frac{dC}{dt},$$

or the ratio of anthropogenic CO₂ emissions $f(t)$ that is not absorbed by land or ocean and thus remains in the atmosphere (“airborne”), converges to a constant (Bennedsen et al., 2019).

To approach this claim with the theory of differential equations, we note the global carbon budget equation, which states that anthropogenic CO₂ emissions are either absorbed by land, ocean, or the atmosphere. The instantaneous change in atmospheric CO₂ is thus given by

$$\frac{dC(t)}{dt} = f(t) - \left(\frac{1}{\tau_{oc}} + \frac{1}{\tau_{ld}} \right) C(t).$$

With $x(t) := C(t)$ and $c := \tau_{oc}^{-1} + \tau_{ld}^{-1} > 0$ obtains

$$\dot{x}(t) + cx(t) = f(t).$$

Multiplying with the integrating factor e^{ct} yields

$$\dot{x}(t)e^{ct} + ce^{ct}x(t) = f(t)e^{ct} = \frac{d}{dt}(xe^{ct}).$$

Integrating, we obtain

$$xe^{ct} = x_0 + \int_0^t f(s)e^{cs} ds,$$

or,

$$x = x_0e^{-ct} + e^{-ct} \int_0^t f(s)e^{cs} ds.$$

Plugging in the assumption for emissions that $f(t) = f_0e^{dt}$, we get

$$\begin{aligned} x &= x_0e^{-ct} + e^{-ct} f_0 \int_0^t e^{(c+d)s} ds \\ &= x_0e^{-ct} + f_0e^{-ct} \left[\frac{1}{c+d} e^{(c+d)s} \right]_0^t \\ &= x_0e^{-ct} + f_0e^{-ct} \left(\frac{1}{c+d} e^{(c+d)t} - \frac{1}{c+d} \right). \end{aligned}$$

In order to calculate the airborne fraction, we need a solution for x/f :

$$\begin{aligned} \frac{x(t)}{f(t)} &= f_0^{-1} e^{-dt} \left(x_0e^{-ct} + f_0e^{-ct} \left(\frac{1}{c+d} e^{(c+d)t} - \frac{1}{c+d} \right) \right) \\ &= \frac{x_0}{f_0} e^{-(c+d)t} + \frac{1}{c+d} - \frac{1}{c+d} e^{-(c+d)t} \\ &= \left(\frac{x_0}{f_0} - \frac{1}{c+d} \right) e^{-(c+d)t} + \frac{1}{c+d}. \end{aligned}$$

We arrive at

$$\begin{aligned}
 AF = \frac{\dot{x}(t)}{f(t)} &= \frac{f(t) - cx(t)}{f(t)} = 1 - c \frac{x(t)}{f(t)} \\
 &= 1 - c \left(\frac{x_0}{f_0} - \frac{1}{c+d} \right) e^{-(c+d)t} - \frac{c}{c+d}.
 \end{aligned}$$

As $t \rightarrow \infty$, the airborne fraction converges to

$$AF = \frac{\dot{x}(t)}{f(t)} \rightarrow_{t \rightarrow \infty} 1 - \frac{c}{c+d} = \frac{d}{c+d}.$$

4.4 The Ramsey model as a recurring example throughout a course

I use the Ramsey model as a recurrent example throughout the third-semester undergraduate course based on Sydsæter et al. (2008), since it touches on many elements that we cover. The sequence of examples from the Ramsey model culminates in the optimization problem for the household, for which control theory is necessary. The entire course also culminates in many respects in the chapter on control theory, which draws on elements from linear algebra, multivariate calculus, static optimization, and differential equations.

4.4.1 CRR or CIES utility functions

This is a preliminary exercise that provides an insight that is needed later in the solution of the Ramsey model, but of course it is also of independent interest. Solving the exercise requires only methods from first-order ordinary differential equations. It is also the only example for a differential equation where the indicator variable is not time in my course.

Let $U : x \mapsto U(x) \in \mathbb{R}$ be a twice continuously differentiable utility function. The elasticity of the first derivative U' with respect to the good x ,

$$-\epsilon_x(U') = - \frac{dU'(x)}{U'(x)} \bigg/ \frac{dx}{x} = - \frac{dU'(x)}{dx} \frac{x}{U'(x)} = - \frac{U''(x)x}{U'(x)}$$

is called the *Arrow-Pratt measure* of relative risk aversion. By requiring constant relative risk aversion (CRR), or constant (intertemporal) elasticity of substitution (C(IES)), one poses a second-order differential equation for U :

$$-\frac{U''(x)x}{U'(x)} \equiv \theta \iff U''(x) = -\frac{\theta}{x} U'(x), \theta > 0.$$

Define the substitution $v(x) := U'(x)$. Then,

$$v'(x) = -\frac{\theta}{x} v(x)$$

Solve this first-order DE for v and then integrate v to obtain U . Distinguish the cases $\theta \neq 1$ and $\theta = 1$. The resulting class of functions is the family of CRR or C(I)ES utility functions. The higher θ , the faster the relative decline in $U'(x)$ in response to increases in x . That is, large increases in x are less welcome and large decreases in x are more avoided than with a lower value of θ .

We obtain

$$\begin{aligned} v'(x) &= -\frac{\theta}{x} v(x), \\ \int \frac{v'(x)}{v(x)} dx &= \int \frac{1}{v(x)} dv = - \int \frac{\theta}{x} dx, \\ \log v(x) &= -\theta(\log x + c), \quad c \in \mathbb{R}, \\ v(x) &= e^{-\theta c} x^{-\theta} =: k_0 x^{-\theta}. \end{aligned}$$

Then,

$$\begin{aligned} U(x) &= \int U'(x) dx = \int v(x) dx \\ &= \int k_0 x^{-\theta} dx = \frac{x^{1-\theta}}{1-\theta} k_0 + k_1, \quad k_1 \in \mathbb{R}, \end{aligned}$$

for $\theta \neq 1$. For $\theta = 1$,

$$\int k_0 x^{-1} dx = k_0 \log x + k_1.$$

This is the family of CRR or C(I)ES utility functions. The higher θ , the faster the relative decline in $U'(x)$ in response to increases in x . That is, large increases in x are less welcome and large decreases in x are more avoided than with a lower value of θ .

4.4.2 Linearization around steady-state as example for Taylor expansion

The first point of contact of the students with the Ramsey model is in multivariate calculus when we cover the Taylor expansion. I introduce the central system of differential equations that describes capital and consumption dynamics in the model, even though the students have not seen differential equations at this point. Since the task is to understand the linearization of the right-hand sides of the differential equations around the steady state and not to solve the equations, this is not material. I usually spend a few words explaining what the idea of differential equations is and that we are going to come back to this later in the course. I recurrently use Cobb-Douglas production and CES utility functions as examples throughout the course, so the students are familiar with them at this point. They also take Macroeconomics I in the same semester. I use three slides to introduce the following material.

In macroeconomics, the Ramsey model is a key model to describe consumption and production in a stylized economy (Ramsey, 1928; Koopmans, 1963; Cass, 1965),

with textbook treatments, for example, in Barro and Sala-i Martin (2004) and Romer (2006).

In equilibrium, the economy is described by a system of non-linear ordinary differential equations for capital per capita $k(t)$ and consumption per capita $c(t)$, as functions of time.

$$\begin{aligned}\frac{d}{dt} \log k(t) &= Ak^{\alpha-1} - \frac{c}{k} - (\delta + \alpha_L + \alpha_T), \\ \frac{d}{dt} \log c(t) &= \frac{1}{\theta} \left(\alpha Ak^{\alpha-1} - (\delta + \rho + \theta \alpha_T) \right).\end{aligned}$$

- $A, \alpha > 0$ are the parameters of a Cobb-Douglas production function,
- $\theta > 0$ is the parameter of a CES utility function,
- $\delta > 0$ is the depreciation of the capital stock,
- $\alpha_L > 0$ is the growth rate of the labor force,
- $\alpha_T > 0$ is the rate of technological progress,
- $\rho > 0$ is the rate of time preference.

The *steady state* $(k(t), c(t)) \equiv (k^*, c^*)$, where neither capital nor consumption change through time, is defined by

$$\frac{d}{dt} \log k(t) = \frac{d}{dt} \log c(t) = 0.$$

The fact that the system of differential equations is non-linear is often cumbersome, for example in numerical analyses. For many purposes, the system is linearized around the steady state. That is, the function

$$\begin{aligned}f : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ \begin{bmatrix} \log k \\ \log c \end{bmatrix} &\longmapsto \begin{bmatrix} \frac{d}{dt} \log k \\ \frac{d}{dt} \log c \end{bmatrix},\end{aligned}$$

is expanded around $(\log k^*, \log c^*)$ to first order.

The resulting first-order Taylor expansion is a linear system of differential equations:

$$\begin{bmatrix} \frac{d}{dt} \log k \\ \frac{d}{dt} \log c \end{bmatrix} = \begin{bmatrix} \rho - \alpha_L - (1 - \theta)\alpha_T & \delta + \alpha_L + \alpha_T - \frac{1}{\alpha}(\delta + \rho + \theta\alpha_T) \\ \frac{\alpha-1}{\theta}(\delta + \rho + \theta\alpha_T) & 0 \end{bmatrix} \begin{bmatrix} \log \frac{k}{k^*} \\ \log \frac{c}{c^*} \end{bmatrix}.$$

To fill in the blanks in this example is the following exercise in a homework set.

1. Show that the Cobb-Douglas production function $Y = AK^\alpha L^{1-\alpha}$ can be written in per-capita form

$$y = Ak^\alpha,$$

where $y = Y/L$ and $k = K/L$.

2. Show that the steady-state values for capital and consumption in the Ramsey model are given by

$$k^* = \left[\frac{1}{\alpha A} (\delta + \rho + \theta \alpha_T) \right]^{\frac{1}{\alpha-1}},$$

$$c^* = A(k^*)^\alpha - (\delta + \alpha_L + \alpha_T)k^*.$$

Set $\frac{d}{dt} \log k(t) = \frac{d}{dt} \log c(t) = 0$ and solve for c og k .

3. Show that the Jacobian of

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2,$$

$$\begin{bmatrix} \log k \\ \log c \end{bmatrix} \mapsto \begin{bmatrix} Ak^{\alpha-1} - \frac{c}{k} - (\delta + \alpha_L + \alpha_T) \\ \frac{1}{\theta} (\alpha Ak^{\alpha-1} - (\delta + \rho + \theta \alpha_T)) \end{bmatrix},$$

evaluated in $(\log k^*, \log c^*)$ is given by

$$J_f(\log k^*, \log c^*) = \begin{bmatrix} \rho - \alpha_L - (1 - \theta)\alpha_T & \delta + \alpha_L + \alpha_T - \frac{1}{\alpha}(\delta + \rho + \theta \alpha_T) \\ \frac{\alpha-1}{\theta}(\delta + \rho + \theta \alpha_T) & 0 \end{bmatrix}.$$

4. Show that

$$f(\log k^*, \log c^*) = (0, 0)^T.$$

Hint: Rewrite the function f to

$$\begin{bmatrix} \log k \\ \log c \end{bmatrix} \mapsto \begin{bmatrix} Ae^{(\alpha-1)\log k} - e^{\log \frac{c}{k}} - (\delta + \alpha_L + \alpha_T) \\ \frac{1}{\theta} (\alpha Ae^{(\alpha-1)\log k} - (\delta + \rho + \theta \alpha_T)) \end{bmatrix}$$

with Jacobian

$$J_f = \begin{bmatrix} A(\alpha-1)k^{\alpha-1} + \frac{c}{k} - \frac{c}{k} \\ \frac{\alpha(\alpha-1)}{\theta} Ak^{\alpha-1} & 0 \end{bmatrix}.$$

4.4.3 Understanding the household's optimization problem

When we cover first-order linear ordinary differential equations (Chapter 5 in Sydsæter et al. (2008)), I spend three slides on explaining the household's optimization problem. The students cannot solve it yet, but they can understand the objects that figure in the problem.

In the Ramsey model introduced in Chapter 2, define the following objects.

- $L(t) = e^{\alpha_L t}$ work force per representative household
 $\alpha_L > 0$ population growth rate
 $C(t)$ aggregate consumption of the representative household
 $c(t) = \frac{C(t)}{L(t)}$ per-capita consumption
 $u(x)$ CRRA, CES utility function
 $\rho > 0$ rate of time preference
 $A(t)$ total assets of the representative household
 $a(t) = \frac{A(t)}{L(t)}$ per-capita assets
 $r(t)$ returns on assets
 $w(t)$ wage rate per unit of labor

Assets can be ownership of capital or loans to firms; both earn the return rate $r(t)$. Negative values of a are possible, they represent debt to firms. The instantaneous utility function is given by

$$u(x) = \frac{x^{1-\theta} - 1}{1-\theta}.$$

The utility of the representative household is then given by integrating over the individual utility times the number of individuals times a discount factor to account for time preference. Assuming the planning horizon $[t_0, t]$, the objective function of the representative household is

$$J(t, a, c) = \int_{t_0}^t u(c(s))L(s)e^{-\rho s} ds.$$

This objective function is maximized subject to a constraint: The present value of all consumption during the planning interval plus the present value of any assets left at time t must equal the present value of all income plus the assets owned at the beginning of the period, t_0 .

The instantaneous change in the total asset position $A(t)$ of the representative household is given by returns on the total asset position, $r(t)A(t)$, plus total wages per household, $w(t)L(t)$, minus household consumption $C(t)$:

$$\dot{A}(t) = r(t)A(t) + w(t)L(t) - C(t),$$

or in per-capita terms,

$$\dot{a}(t) = r(t)a(t) + w(t) - c(t) - \alpha_L a(t).$$

(Show that $\dot{a}(t) = \dot{A}(t) - \alpha_L a(t)$.) The solution for $a(t)$ is obtained by variation of constants as

$$a(t) = a(t_0)e^{\int_{t_0}^t (r(s) - \alpha_L) ds} + e^{\int_{t_0}^t (r(s) - \alpha_L) ds} \int_{t_0}^t (w(s) - c(s)) e^{-\int_{t_0}^s (r(\tau) - \alpha_L) d\tau} ds,$$

or

$$a(t)e^{-\int_{t_0}^t (r(s) - \alpha_L) ds} + \int_{t_0}^t c(s)e^{-\int_{t_0}^s (r(\tau) - \alpha_L) d\tau} ds = a(t_0) + \int_{t_0}^t w(s)e^{-\int_{t_0}^s (r(\tau) - \alpha_L) d\tau} ds.$$

The last equation captures the idea of the constraint explained verbally above.

4.4.4 Solving the linearized system of differential equations

When we get to second-order linear differential equations (Chapter 6 in Sydsæter et al. (2008)), we have the tools to solve the linearized system we found in the first exercise:

In the Ramsey model, we found a linearization of the logarithmic capital stock $\log k(t)$ and the logarithmic consumption $\log c(t)$ around their steady-state values $(\log k^*, \log c^*)$ as

$$\dot{x}(t) = Ax(t) + b,$$

where

$$x(t) = \begin{bmatrix} \log k(t) \\ \log c(t) \end{bmatrix},$$

and

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

with entries

$$\begin{aligned} a_{11} &= \rho - \alpha_L - (1 - \theta)\alpha_T, \\ a_{12} &= \delta + \alpha_L + \alpha_T - \frac{1}{\alpha}(\delta + \rho + \theta\alpha_T), \\ a_{21} &= \frac{\alpha - 1}{\theta}(\delta + \rho + \theta\alpha_T), \\ a_{22} &= 0. \end{aligned}$$

The inhomogeneous term is

$$b = -A \begin{bmatrix} \log k^* \\ \log c^* \end{bmatrix},$$

where

$$\begin{aligned} k^* &= \left[\frac{1}{\alpha A}(\delta + \rho + \theta\alpha_T) \right]^{\frac{1}{\alpha-1}}, \\ c^* &= A(k^*)^\alpha - (\delta + \alpha_L + \alpha_T)k^*. \end{aligned}$$

Exercise: Determine the eigenvalues $\lambda_{1,2}$ and eigenvectors $v_{1,2}$ of A (as functions of the a_{ij}). Show that the matrix A is diagonalizable if and only if

$$a_{11}^2 \neq -4a_{12}a_{21}.$$

The solutions have the form

$$[v_1 e^{\lambda_1 t}, v_2 e^{\lambda_2 t}].$$

The inhomogeneous equation

$$\dot{x}(t) = Ax(t) + b,$$

where $b = -Ax^*$ og $x^* = (\log k^*, \log c^*)$, is solved with the substitution $z = \log k - \log k^*$ and $w = \log c - \log c^*$.

Eigenvalues and -vectors of A are given by

$$\begin{aligned}\lambda_1 &= \frac{a_{11}}{2} + \sqrt{\frac{a_{11}^2}{4} + a_{12}a_{21}}, & v_1 &= \begin{bmatrix} -a_{12} \\ \frac{a_{11}}{2} - \sqrt{\frac{a_{11}^2}{4} + a_{12}a_{21}} \end{bmatrix}, \\ \lambda_2 &= \frac{a_{11}}{2} - \sqrt{\frac{a_{11}^2}{4} + a_{12}a_{21}}, & v_2 &= \begin{bmatrix} -a_{12} \\ \frac{a_{11}}{2} + \sqrt{\frac{a_{11}^2}{4} + a_{12}a_{21}} \end{bmatrix}.\end{aligned}$$

4.4.5 Solving the optimization problems and closing the model

When we have covered dynamic optimization problems with infinite horizon in the chapter on control theory (Chapter 9 in Sydsæter et al. (2008)), I present the students with this “soft” exercise. In the oral exam, the students draw a card with a subject out of a dozen or so, covering different parts of the course. They prepare an 8-minute presentation of each subject, so that they have a degree of control. I agree in advance with the students that I will only talk about the Ramsey model, if they choose it in their presentation when they draw a subject in the chapter on systems of differential equations or in control theory. In the ten years that I have been teaching this course, no student has chosen to do so, even though some have expressed appreciation of the coverage of the model.

This exercise comprises understanding the text and following the steps to arrive at the displayed equations. Throughout this class, we have used elements of the Ramsey model in macroeconomics as a motivation. We are now in a position to derive the first-order conditions for optimality in this model, which are a system of differential equations. The notation used here resembles the one in Barro and Sala-i-Martin: *Economic Growth*, 2nd ed. 2004.

The general structure of the model is that firms and households (more precisely, the representative firm and the representative household) have independent optimization problems. Firms maximize profits with respect to capital input. Profits are price times output minus costs, which are returns on capital to be paid to the owner of the capital, capital depreciation, and wages for labor employed. Output is a (Cobb-Douglas production) function of capital and labor. Firms take the price for the final product as datum, and it is customarily normalized to one. They also take the return rate on capital and the wage rate per unit of labor as data, meaning that they cannot

choose these values. It turns out that the firm's optimization problem can be solved by means of calculus and need not be cast as an optimal control problem.

Households maximize utility for their working members over an infinite time horizon. Utility is a function of consumption, which can be paid for by either wages for labor or by returns on assets that the household owns. Revenue from either sources that are not consumed are saved, build up assets, and are thus the source for future returns. We have studied the dynamics of the asset build-up as an example for a univariate inhomogeneous linear ordinary differential equation and determined the solution in an example in the chapter on first-order ordinary differential equations. The household's optimization problem is an optimal control problem. The objective function depends on the state variable assets, which in turn depends on the control variable consumption. The differential transition equation for the budget constraint in the example in the chapter on first-order differential equations describes the relation between assets and consumption and constitutes the main constraint of the optimization problem. The household optimization takes wages and returns on assets as data.

The two independent optimization problems of the firms and of the households, respectively, therefore take two variables as data: returns on assets and wages in the case of the household, and returns on capital and wages in the case of the firm. These loose ends are now tied down by imposing equilibrium. The equilibrium conditions "close" the model. The wage variable is obviously the same for both entities, firms and households, and by principles of microeconomics, in equilibrium it must equal the marginal product of labor, which can be determined from the production function. Returns on assets and returns on capital are imposed to coincide by requiring that in equilibrium, the total asset stock equals the total capital stock. This models ownership of the productive capital stock by the households. Note that labor is not a choice variable in the Ramsey model, neither the labor supply by households nor the labor demand by firms. Economists say that labor is determined exogenously, meaning that it is not chosen by any agent inside the model. The labor force simply grows exponentially at a parameter rate. Exogenous parameters that describe agent's behavior, but are not consciously and optimally chosen by the agents, such as the rate of time preference or the parameter of risk aversion in the utility function, are often called "deep parameters."

In the example on the household budget constraint in the Ramsey model in the chapter on first-order differential equations, we introduced the following objects:

$L(t) = e^{\alpha_L t}$	work force per representative household
$\alpha_L > 0$	population growth rate
$C(t)$	aggregate consumption of the representative household
$c(t) = \frac{C(t)}{L(t)}$	per-capita consumption
$u(x)$	$= (x^{1-\theta} - 1)/(1 - \theta)$, CRRA, CES utility function
$\rho > 0$	rate of time preference
$A(t)$	total assets of the representative household
$a(t) = \frac{A(t)}{L(t)}$	per-capita assets
$r(t)$	returns on assets
$w(t)$	wage rate per unit of labor

We will need the following additional objects; most of them are related to the firm's optimization problem.

$K(t)$	capital stock
$\delta > 0$	capital depreciation rate
$Y(t)$	output
$T(t) = e^{\alpha_T t}$	technology
$\alpha_T > 0$	rate of technological progress
$L(t)$	$(= L(t)T(t) = L(t)e^{\alpha_T t})$ effective labor

It is common to work in *units of effective labor*. We define

$$y := \frac{Y}{L}, \quad k := \frac{K}{L}, \quad c := \frac{C}{L},$$

as output, capital, and consumption, respectively, in units of effective labor. Note that $L(t) = e^{(\alpha_L + \alpha_T)t}$, and that the initial labor force is normalized to one unit. Then,

$$c = \frac{C}{L} = \frac{C}{e^{\alpha_L t}}, \quad c = \frac{C}{e^{\alpha_L + \alpha_T t}} = c e^{-\alpha_T t}.$$

Analogously, we define $k = K/L$, so that $k = k e^{-\alpha_T t}$.

We begin with the firm's problem. The production function is of the Cobb-Douglas type, where technology augments labor

$$Y(t) = F(K(t), L(t)) = AL(t)^{1-\alpha} K(t)^\alpha = A(L(t)T(t))^{1-\alpha} K(t)^\alpha,$$

so that the total technological impact on labor productivity is $A(T(t))^{1-\alpha}$. In units of effective labor,

$$y = \frac{Y}{L} = F(k, 1) =: f(k) = Ak^\alpha.$$

Firms pay return on capital (to their owners, who are the households), lose depreciation on capital stock, and pay wages. Remember that the price of the final good is normalized to one. We then have the profit function

$$\begin{aligned}
\Pi &= F(K, L) - (r + \delta)K - wL \\
&= Lf(k) - (r + \delta)Lk - wLe^{-\alpha_T t} \\
&= L(f(k) - (r + \delta)k - we^{-\alpha_T t}).
\end{aligned}$$

Maximizing this profit function over any time interval is equal to maximizing at every single point in time since there is no state variable involved that could transfer residual amounts to the future or borrow from the future. Differentiation with respect to k and setting equal to zero results in the first-order condition

$$r(t) = f'(k(t)) - \delta, \quad (1)$$

which determines the return on productive capital intuitively as the marginal product minus depreciation.

Turning to the household's problem, there is a state variable, assets, which can delay consumption, and thus utility, to the future or borrow consumption, and thus utility, from the future. (The case $a(t) < 0$ is allowed and interpreted as debt of households to firms.) The household's problem is a proper optimal control problem. The problem is

$$\begin{aligned}
\max_{\mathbf{c}} J(t, a, \mathbf{c}) &= \int_{t_0}^{\infty} u(\mathbf{c}(s))L(s)e^{-\rho s} ds, \\
\text{such that } \dot{\mathbf{a}}(t) &= r(t)\mathbf{a}(t) + w(t) - \mathbf{c}(t) - \alpha_L \mathbf{a}(t), \\
\mathbf{a}(0) &= \mathbf{a}_0.
\end{aligned}$$

Assets per capita increase by the returns r on assets, and by the difference of wages and consumption, and finally by the growth rate of the labor force (the denominator of \mathbf{a}). There is an additional constraint that rules out Ponzi schemes (chain letters, meaning constant borrowing at a faster rate than the interest rate, so that at the end of the period, a very large debt position remains):

$$\lim_{t \rightarrow \infty} \mathbf{a}(t) e^{-\int_{t_0}^t (r(s) - \alpha_L) ds} \geq 0.$$

It will turn out later that this condition is met by the transversality condition. The Hamiltonian is given by

$$H(t, \mathbf{a}, \mathbf{c}, \nu) = u(\mathbf{c})e^{-(\rho - \alpha_L)t} + \nu [w + (r - \alpha_L)\mathbf{a} - \mathbf{c}].$$

The first-order conditions according to the maximum principle are:

$$\frac{\partial H}{\partial \mathbf{c}} = e^{-(\rho - \alpha_L)t} u'(\mathbf{c}) - \nu = 0, \quad (2)$$

$$\dot{\nu} = -\frac{\partial H}{\partial \mathbf{a}} = -\nu(r - \alpha_L), \quad (3)$$

$$\lim_{t \rightarrow \infty} \nu(t)\mathbf{a}(t) = 0. \quad (4)$$

Differentiate (2) with respect to time, plug into (3), and remember from the exercise on constant relative risk aversion that

$$-\frac{u''(c)c}{u'(c)} = \theta,$$

then

$$\frac{\dot{c}}{c} = \frac{1}{\theta}(r(t) - \rho). \quad (5)$$

From (3), we have

$$v(t) = v(0)e^{-\int_0^t (r(s) - \alpha_L) ds},$$

so that the transversality condition reads

$$\lim_{t \rightarrow \infty} a(t)e^{-\int_0^t (r(s) - \alpha_L) ds} = 0,$$

which will ensure meeting the no-Ponzi condition, which states that this object must be positive or zero. Households do not want to have positive assets left over at the end of the total period, that is why strictly positive values are ruled out. If positive assets were left at the end of the period in the optimum, their shadow price v would have to be zero. As long as the shadow price is positive, households do not want to waste assets at the end. In this sense, the transversality condition is the continuous-time equivalent of complementary slack in the Kuhn-Tucker conditions.

We now “close” the model by imposing equilibrium conditions. One is the microeconomic insight that in equilibrium, the wage rate $w(t)$ must equal the marginal product of labor, which we obtain from the production function:

$$\begin{aligned} Y &= F(K, L) = Lf(k), \\ \frac{\partial F}{\partial L} &= \frac{\partial L}{\partial L} f(k) + Lf'(k) \frac{\partial K}{\partial L} \frac{1}{L} = e^{\alpha_T t} (f(k) - f'(k)k) = w(t). \end{aligned}$$

The second equilibrium condition states that assets equal capital $a(t) = k(t)$ for all t , and thus

$$\dot{k} = \dot{a} = (r - \alpha_L)a + w - c.$$

By definition of k and k , we have

$$\begin{aligned} \dot{k} &= \dot{k}e^{-\alpha_T t} - k\alpha_T e^{-\alpha_T t}, \\ &= ((r - \alpha_L)a + w - c)e^{-\alpha_T t} - k\alpha_T e^{-\alpha_T t}, \\ &= f(k) - c - (\delta + \alpha_L + \alpha_T)k, \end{aligned} \quad (6)$$

where the last equation uses the equilibrium condition for w , the first-order condition (1) from the firm's problem, and $a = k$.

From the household's problem, we distilled the condition (5). Verifying that

$$\frac{\dot{c}}{c} = \frac{\dot{c}}{c} - \alpha_T,$$

and using (1) from the firm's problem, we obtain

$$\frac{\dot{c}}{c} = \frac{1}{\theta}(f'(k) - \delta - \rho - \theta\alpha_T). \quad (7)$$

Inserting the Cobb-Douglas specification $f(k) = ak^\alpha$, Equations (6) and (7) result in the system of non-linear ordinary differential equations stated in the example in our first encounter with the model:

$$\begin{aligned} \frac{d}{dt} \log k(t) = \frac{\dot{k}}{k} &= Ak^{\alpha-1} - \frac{c}{k} - (\delta + \alpha_L + \alpha_T), \\ \frac{d}{dt} \log c(t) = \frac{\dot{c}}{c} &= \frac{1}{\theta} \left(\alpha Ak^{\alpha-1} - (\delta + \rho + \theta\alpha_T) \right). \end{aligned}$$

5 Conclusion

In this chapter, I have reflected on the teaching of mathematics for students of economics. I outlined an ideal course sequence if the objective is to enable the student to understand (undergraduate) and contribute to (graduate) the economic literature. Mathematical modeling in economics is briefly discussed. Three years of experience with teaching Mathematics for Economists online were summarized. The second half of the chapter offered a repository of examples and exercises for instructors of Mathematics for Economists to use freely. They covered a large range of subjects in linear algebra, analysis, differential equations, and dynamic optimization. The Ramsey model figured prominently among these examples.

Acknowledgements I would like to thank my students and teaching assistants in 2622 Mathematics for Economists over the years. The former TAs Rebekka Jørgensen, Marie Møller Schmidt, and Hannah Segato have read and commented on a draft of this paper. I am grateful for careful reading and comments from Ludwig Arnold, Mikkel Børgesen, and Jan Magnus.

Competing Interests The author has no conflicts of interest to declare that are relevant to the content of this chapter.

References

- Arrow KJ (1963) Social Choice & Individual Values, 2nd edn. Cowles Foundation
- Arrow KJ, Intriligator MD (1981) Handbook of Mathematical Economics Vol. I. North Holland
- Arrow KJ, Intriligator MD (1982) Handbook of Mathematical Economics Vol. II. North Holland

- Arrow KJ, Intriligator MD (1986) Handbook of Mathematical Economics Vol. III. North Holland
- Barro R, Sala-i Martin X (2004) Economic Growth, 2nd edn. MIT Press
- Bennedsen M, Hillebrand E, Koopman SJ (2019) Trend analysis of the airborne fraction and sink rate of anthropogenically released CO₂. *Biogeosciences* 16(18):3651–3663
- Bennedsen M, Hillebrand E, Koopman SJ (2023) A multivariate dynamic statistical model of the global carbon budget 1959–2020. *Journal of the Royal Statistical Society Series A: Statistics in Society* 186(1):20–42
- Box G (1979) Robustness in the strategy of scientific model building. In: Launer R, Wilkinson G (eds) *Robustness in Statistics*, Academic Press, pp 201–236, DOI <https://doi.org/10.1016/B978-0-12-438150-6.50018-2>, URL <https://www.sciencedirect.com/science/article/pii/B9780124381506500182>
- Boyd S, Vandenberghe L (2004) *Convex Optimization*. Cambridge University Press
- Cacault MP, Hildebrand C, Laurent-Lucchetti J, Pellizzari M (2021) Distance learning in higher education: evidence from a randomized experiment. *Journal of the European Economic Association* 19(4):2322–2372
- Cass D (1965) Optimum growth in an aggregative model of capital accumulation. *The Review of Economic Studies* 32(3):233–240
- Chiang AC (1984) *Fundamental Methods of Mathematical Economics*, 3rd edn. McGraw-Hill
- Corbae D, Stinchcombe MB, Zeman J (2009) *An Introduction to Mathematical Analysis for Economic Theory and Econometrics*. Princeton University Press
- Debreu G (1959) *Theory of Value*. Cowles Foundation
- Dhrymes PJ (2013) *Mathematics for Econometrics*, 4th edn. Springer
- Dixit AK (1990) *Optimization in Economic Theory*. Oxford University Press
- Hoy M, Livernois J, McKenna C, Rees R, Stengos T (2001) *Mathematics for Economics*, 2nd edn. MIT Press
- Intriligator MD (2002) *Mathematical Optimization and Economic Theory*. SIAM
- Johnsonbaugh R, Pfaffenberger W (2010) *Foundations of Mathematical Analysis*. Dover
- Klein MW (2002) *Mathematical Methods for Economics*, 2nd edn. Addison Wesley
- Koopmans TC (1963) On the concept of optimal economic growth. *Cowles Foundation Discussion Papers* (392)
- Leijonhufvud A (1973) Life Among the Econ. *Western Economic Journal* (now *Economic Inquiry*) 11:327–337
- Luenberger DG (1969) *Optimization by Vector Space Methods*. Wiley
- Nelsen RB (2000) *Proofs Without Words II: More exercises in visual thinking*. The Mathematical Association of America
- von Neumann J (1995) Method in the physical sciences. In: Bródy F, Vámos T (eds) *The Neumann Compendium*, World Scientific, p 628
- Novshek W (1993) *Mathematics for Economists*. Academic Press
- Ok EA (2007) *Real Analysis with Economic Applications*. Princeton University Press

- Ramsey FP (1928) A mathematical theory of saving. *The Economic Journal* 38(152):543–559
- Resnick SI (2014) *A Probability Path*. Springer
- Romer D (2006) *Advanced Macroeconomics*, 3rd edn. McGraw-Hill
- Rudin W (1976) *Principles of Mathematical Analysis*, 3rd edn. McGraw-Hill
- Schindler S (2018) *Theoretical Virtues in Science: Uncovering Reality through Theory*. Cambridge University Press
- Seierstad A, Sydsæter K (1987) *Optimal Control Theory with Economic Applications*. North Holland
- Simon CP, Blume L (1994) *Mathematics for Economists*. Norton
- Stigum BP (1990) *Toward a Formal Science of Economics*. MIT Press
- Strang G (2023) *Introduction to Linear Algebra*, sixth edn. Cambridge University Press
- Sydsæter K, Hammond P, Seierstad A, Strøm A (2008) *Further Mathematics for Economic Analysis*, 2nd edn. Pearson
- Sydsæter K, Hammond P, Strøm A, Carvajal A (2022) *Essential Mathematics for Economic Analysis*, sixth edn. Pearson
- Thompson E (2022) *Escape from Model Land: How mathematical models can lead us astray and what we can do about it*. Basic Books
- Weintraub ER (1982) *Mathematics for economists: An integrated approach*. Cambridge University Press