Modelling and Simulation ESS101
26 October 2022, Final exam

This exam contains 10 pages (including this cover page) and 5 problems.

You are allowed to use the following material:

- *Modelling And Simulation, Lecture notes for the Chalmers course ESS101*, by S. Gros (with brief annotations, but cannot contain solutions to the exercises or previous exams)
- *Mathematics Handbook* (Beta)
- *Physics Handbook*
- Chalmers approved calculator
- Formula sheet, appended to the exam.

– Organize your work in a reasonably neat and coherent way. Work scattered all over the page without a clear ordering may receive less credit.

– Mysterious or unsupported answers will not receive credit, but an incorrect answer supported by substantially correct calculations and explanations will receive partial credit.

– None of the proposed questions require extremely long computations. If you get caught in endless algebra, you have probably missed the simple way of doing it.

– The passing grade will be given at 20 points, grade 4 at 27 and the top grade at 34 points.

Problem	Points	Score
1	8	
2	8	
3	11	
4	6	
5	7	
Total:	40	

Best of luck to all !!

Examiner: Yasemin Bekiroglu, +46 70 148 72 71
Teaching Assistant: Remi Lacombe, +46 70 716 34 71

1. (a) (8 points) Consider a 3D roller-coaster created by a mass m sliding on a surface described by the (scalar-valued) constraint equation $c(\mathbf{p}) = 0$ (the exact constraint definition is not needed), where $\mathbf{p} \in \mathbb{R}^3$ is the usual cartesian coordinate vector. The mass is affected by gravity and subject to a friction force given by $\mathbf{F} = -m\gamma\dot{\mathbf{p}}$.
    1. (4 points) Determine the Lagrange function for the system.
    2. (4 points) Derive the Euler-Lagrange equations.

Solution:

(a) 1. The kinetic and potential energies of the system can be written as ($\mathbf{q} = \mathbf{p}$):

$$T = \frac{1}{2}m\dot{\mathbf{p}}^\top\dot{\mathbf{p}}, \qquad V = mg\mathbf{e}^\top\mathbf{p} \tag{1}$$

where $\mathbf{e}^\top = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$. The Lagrange function then reads as:

$$\mathcal{L} = T - V - zc = \frac{1}{2}m\dot{\mathbf{p}}^\top\dot{\mathbf{p}} - mg\mathbf{e}^\top\mathbf{p} - zc\left(\mathbf{p}\right) \tag{2}$$

2. The dynamics are constructed using:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{q}}}^\top = m\ddot{\mathbf{p}}, \qquad \frac{\partial\mathcal{L}}{\partial\mathbf{q}}^\top = -mg\mathbf{e} - z\frac{\partial c}{\partial\mathbf{p}}^\top \tag{3}$$

Since we work in cartesian coordinates, the generalized force attached to the friction is simply $\mathbf{Q} = \mathbf{F}$. The model then follows from Euler-Lagrange's equation:

$$m\ddot{\mathbf{p}} + mg\mathbf{e} + z\frac{\partial c}{\partial\mathbf{p}}^\top = -m\gamma\dot{\mathbf{p}} \tag{4a}$$

$$0 = c(\mathbf{p}) \tag{4b}$$

2. (a) (8 points) Consider the DAE:

$$\dot{\mathbf{x}}_1 = \mathbf{x}_1 + \mathbf{x}_2 + z \tag{5a}$$

$$\dot{\mathbf{x}}_2 = z + u \tag{5b}$$

$$0 = \frac{1}{2}\left(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 1\right) \tag{5c}$$

1. (3 points) Why is it a DAE?
2. (5 points) What is the differential index of (5)?

Solution:

(a) 1. We can answer in a simple or formal way:
- The simple way relies on observing that variable z does not enter as time-differentiated in (5), it is an algebraic variable and therefore
- For the complex and formal way, we observe that (5) is given by the fully implicit differential equation:

$$F\left(\dot{s}, s, u\right) = \begin{bmatrix} \dot{\mathbf{x}}_1 - \mathbf{x}_1 - \mathbf{x}_2 - z \\ \dot{\mathbf{x}}_2 - z - u \\ \frac{1}{2}\left(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 1\right) \end{bmatrix} \tag{6}$$

where $\mathbf{s} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & z \end{bmatrix}^\top$. And observe that matrix

$$\frac{\partial F}{\partial \dot{\mathbf{s}}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{7}$$

is rank-deficient

2. We observe that (5) is a semi-explicit DAE with

$$g(\mathbf{x}, z, u) = \frac{1}{2}\left(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 1\right) \tag{8}$$

and $\frac{\partial g}{\partial z} = 0$, hence it is of index larger than 1. In order to compute precisely the differential index, we need to perform time-differentiations on (5) until it is transformed in an ODE. Because (5a)-(5b) are already ODEs (functions of z), we can leave them alone, and focus on (5c). We then observe that:

$$\frac{\mathrm{d}}{\mathrm{d}t} g(\mathbf{x}, z, u) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\left(\mathbf{x}_1^2 + \mathbf{x}_2^2 - 1\right)\right) = \mathbf{x}_1\dot{\mathbf{x}}_1 + \mathbf{x}_2\dot{\mathbf{x}}_2 \tag{9}$$

Replacing $\dot{\mathbf{x}}_1$, $\dot{\mathbf{x}}_2$ by there expressions from (5a)-(5b), we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}t} g(\mathbf{x}, z, u) = \mathbf{x}_1(\mathbf{x}_1 + \mathbf{x}_2 + z) + \mathbf{x}_2(z + u) \tag{10}$$

which is not yet a differential equations. A second time-derivative delivers:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\mathbf{x}, z, u) = \dot{\mathbf{x}}_1(\mathbf{x}_1 + \mathbf{x}_2 + z) + \dot{\mathbf{x}}_2(z + u) + \mathbf{x}_1(\dot{\mathbf{x}}_1 + \dot{\mathbf{x}}_2 + \dot{z}) + \mathbf{x}_2(\dot{z} + \dot{u}) \tag{11}$$

we then can solve $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\mathbf{x}, z, u) = 0$ for $\dot{z}$:

$$\dot{z} = \frac{-\dot{\mathbf{x}}_1(\mathbf{x}_1 + \mathbf{x}_2 + z) - \dot{\mathbf{x}}_2(z + u) - \mathbf{x}_1\dot{\mathbf{x}}_1 - \mathbf{x}_1\dot{\mathbf{x}}_2 - \mathbf{x}_2\dot{u}}{\mathbf{x}_1 + \mathbf{x}_2} \tag{12}$$

which is an ODE as long as $\mathbf{x}_1 + \mathbf{x}_2 \neq 0$.

3. (a) (5 points) Consider the model:

$$y_k + a_1 y_{k-1} = b_0 u_k + e_k \tag{13}$$

and the associated data $y_{0,\ldots,N}$ and $u_{0,\ldots,N}$ obtained from applying the input sequence $u_{0,\ldots,N}$ to the real system, started with $y_{k<0} = 0$, and e_k is the noise term.

1. (1 point) What kind of model structre is it and why?
2. (2 point) Provide the one-step ahead prediction $\hat{y}_k$ of (13).
3. (2 points) Write the criterion that should be minimized to fit the model to that data, organizing the parameters in a vector $\boldsymbol{\theta} = [a_1 \quad b_0]^T$. Is the least-squares problem linear or nonlinear and why?

(b) (6 points) Consider the model fitting results in Figure 1 and Figure 2.

1. (4 points) In Figure 1, the data pairs are given as $\{[x(i), y(i)]\} = \{[2, 1], [2, 3], [-1, -1]\}$. Using a least squares approach capturing the relationship between x and y, $\hat{\mathbf{y}} = \boldsymbol{\theta}^\top \boldsymbol{\varphi}$ (where $\boldsymbol{\varphi}$ is the *regression vector* holding the *regressors* $(1, x, \ldots)$), find the values in the parameter vector θ that corresponds to the model fit in the figure.

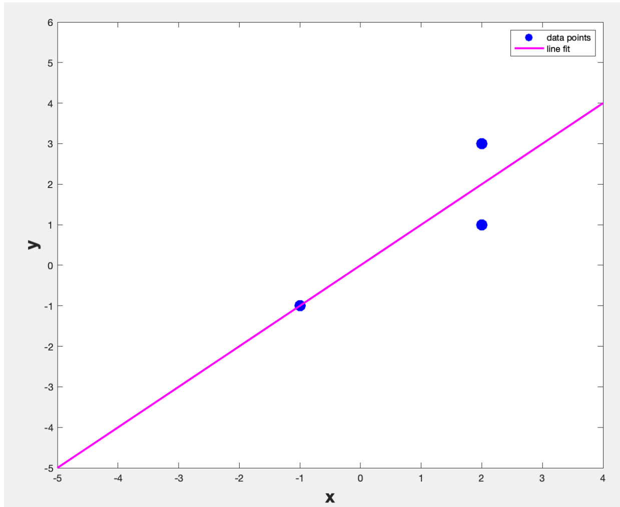

Figure 1: Illustration of three data points and least-squares based model fitting result.

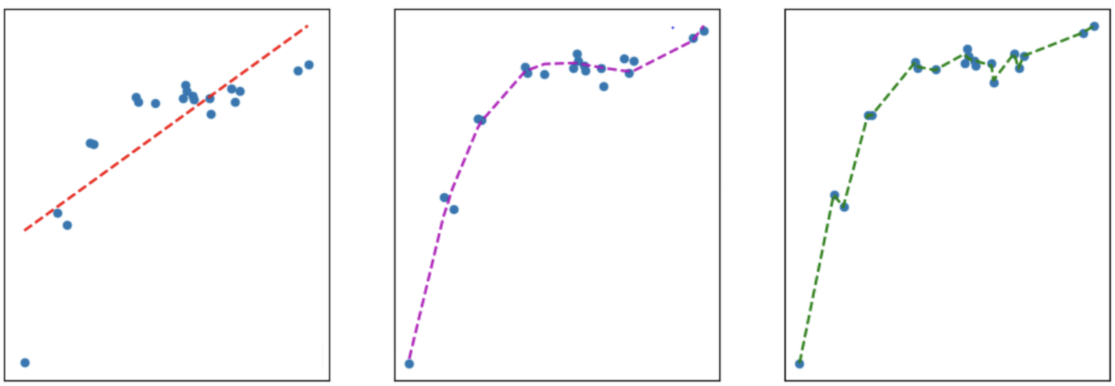

Figure 2: Illustration of raw data points and three different model fitting result.

2. (2 points) Considering the results in Figure 2, comment on which model fit is better and explain why. Comment on issues with worse model fitting results and how to overcome them.

Solution:

(a) 1. ARX.

2. The one-step ahead predictor reads as:

$$\hat{y}_k = -a_1 y_{k-1} + b_0 u_k \tag{14}$$

3. The mismatch between the data and the predictor is given by:

$$e_k = \hat{y}_k - y_k = -y_k - a_1 y_{k-1} + b_0 u_k = \mathbf{d}_k^\top \boldsymbol{\theta} - y_k \tag{15}$$

where $\mathbf{d}_k = \begin{bmatrix} -y_{k-1} & u_k \end{bmatrix}$.

The least-squares problem for the ARX model (13) reads as:

$$\min_{\boldsymbol{\theta}} \quad \frac{1}{2}\|D\boldsymbol{\theta} - \mathbf{y}\|^2 \tag{16}$$

where

$$D = \begin{bmatrix} \mathbf{d}_0^\top \\ \vdots \\ \mathbf{d}_N^\top \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix} \tag{17}$$

The least-squares problem is linear because the regressor $D\boldsymbol{\theta}$ is linear in θ.

(b) 1. Using the data vectors $x = [2, 2, -1]$ and $y = [1, 3, -1]$ in the least-squares estimate where $k = 1$ for fitting a line:

$$\hat{\boldsymbol{\theta}}_N = R_N^{-1} f_N = \Big(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^\top(i)\Big)^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)y(i)$$

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^\top(i) = \frac{1}{N}\begin{bmatrix} N & \sum_{i=1}^{N} x(i) & \dots & \sum_{i=1}^{N} x^k(i) \\ \sum_{i=1}^{N} x(i) & \sum_{i=1}^{N} x^2(i) & \dots & \sum_{i=1}^{N} x^{k+1}(i) \\ . & . & . & \\ . & . & . & \\ . & . & . & \\ \sum_{i=1}^{N} x^k(i) & \sum_{i=1}^{N} x^{k+1}(i) & \dots & \sum_{i=1}^{N} x^{2*k}(i) \end{bmatrix} \tag{18}$$

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)y(i) = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} y(i) \\ \sum_{i=1}^{N} x(i)y(i) \\ . \\ . \\ . \\ \sum_{i=1}^{N} x^k(i)y(i) \end{bmatrix} \tag{19}$$

The resulting parameter estimate $\hat{\boldsymbol{\theta}}_N = [0, 1]^T$, that corresponds to $a = 0$, $b = 1$ in $\hat{\mathbf{y}} = a + bx = \boldsymbol{\theta}^\top\boldsymbol{\varphi}$.

2. The best fit is the second one as the first one has larger error and the last one overfits to the data that leads to poor generalization. Increasing model order solves the problem with the first fitting result and overfitting can be overcome with cross-validation.

4. (a) (4 points) Given the following Butcher tableau

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{2}{3} & \frac{-1}{3} & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 \\ \hline & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{array}$$

1. (3 points) Write down the corresponding RK method equations.
2. (1 points) Is it an explicit or implicit approach, why? How many stages are there?

(b) (1 point) Consider the following RK method:

$$\begin{aligned}\mathbf{K}_1 &= \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}(t_k)\right)\\ \mathbf{K}_2 &= \mathbf{f}\big(\mathbf{x}_k + \frac{\Delta t}{2}\mathbf{K}_1 + \frac{\Delta t}{2}\mathbf{K}_2, \mathbf{u}(t_k + \Delta t)\big)\\ \mathbf{x}_{k+1} &= \mathbf{x}_k + \frac{\Delta t}{2}\mathbf{K}_1 + \frac{\Delta t}{2}\mathbf{K}_2\end{aligned}$$

Is the RK scheme explicit or implicit (explain your answer)? How many stages are there?

(c) (1 point) Explain in your own words why "collocation $\subset$ IRK", i.e. why all collocation methods are IRK methods, but not all IRK methods are collocation methods.

Solution:

(a) 1.

$$\mathbf{K}_1 = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}(t_k)\right) \tag{20a}$$

$$\mathbf{K}_2 = \mathbf{f}\left(\mathbf{x}_k + \frac{\Delta t}{3}\mathbf{K}_1,\ \mathbf{u}\left(t_k + \frac{\Delta t}{3}\right)\right) \tag{20b}$$

$$\mathbf{K}_3 = \mathbf{f}\left(\mathbf{x}_k - \frac{\Delta t}{3}\mathbf{K}_1 + \Delta t\mathbf{K}_2,\ \mathbf{u}\left(t_k + \frac{2\Delta t}{3}\right)\right) \tag{20c}$$

$$\mathbf{K}_4 = \mathbf{f}\left(\mathbf{x}_k + \Delta t\mathbf{K}_1 - \Delta t\mathbf{K}_2 + \Delta t\mathbf{K}_3,\ \mathbf{u}(t_k + \Delta t)\right) \tag{20d}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta t\left(\frac{1}{8}\mathbf{K}_1 + \frac{3}{8}\mathbf{K}_2 + \frac{3}{8}\mathbf{K}_3 + \frac{1}{8}\mathbf{K}_4\right) \tag{21}$$

2. Explicit, the tableau is lower triangular, Ks can be computed sequentially, there are 4 stages, order is 4.

(b) Implicit, The Butcher array is not lower triangular, we cannot compute Ks sequentially as in the Explicit case, we need to use Newton method. It has 2 stages, order is 2.

(c) We know that all collocation methods are implicit RK methods, as the collocation equations are identical to the IRK equations. That explains the inclusion "collocation $\subset$ IRK". However, not all IRK methods are collocation methods. This fact is easy to argue by observing that a collocation method is eventually entirely described by the "checkpoints" $\tau_{1,\dots,\mathrm{s}}$ used to build the polynomials, i.e. we have s "degrees of freedom" to select the collocation method (note that these checkpoints are themselves entirely defined if one want to achieve the order $o = ss$). On the other hand, the Butcher tableau has $\mathrm{s}^2 + 2\mathrm{s}$ entries (a, b, c), and therefore as many degrees of freedom (even though many choices of Butcher tableau can be useless for the purpose of integrating the ODE). This mismatch in the number of degrees of freedom explains why IRK $\not\subseteq$ collocation.

5. (a) (7 points) Considering Newton method,

1. (1 points) When could Newton be unstable where iterations diverge? How could we potentially overcome this problem?
2. (1 point) When could Newton method fail? Explain.
3. (5 points) Consider the function $\mathbf{f} : \mathbb{R}^2 \mapsto \mathbb{R}^2$, by $f(x,y) = \begin{bmatrix} x^2 + y^2 - 4 \\ 4x^2 - y^2 - 4 \end{bmatrix}$, for which we construct approximate solutions to the equation $f(x,y) = [0,0]^T$ using Newton method and

the initial guess $\mathbf{x} = [x_0, y_0] = [1, 1]$. Calculate the resulting solution from applying Newton iteration only once, i.e. $[x_1, y_1]$.

Solution:

(a)
1. Initial guess away from the solution using Full Newton step. We can address this using reduced Newton steps.
2. When the Jacobian is singular, i.e. the Newton direction is undefined. This is a common failure mode, when tackling very nonlinear functions and starting with a poor initial guess.
3. $\frac{\partial f(x,y)}{\partial \mathbf{x}} = \begin{bmatrix} 2x & 2y \\ 8x & -2y \end{bmatrix}$, $det\partial f(x,y) = -20xy$, $\partial f(x,y)^{-1} = \frac{1}{-20xy}\begin{bmatrix} -2y & -2y \\ -8x & 2x \end{bmatrix} = \begin{bmatrix} \frac{1}{10x} & \frac{1}{10x} \\ \frac{2}{5y} & \frac{-1}{10y} \end{bmatrix}$

   Thus, $\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} - \partial f(x_{k-1}, y_{k-1})^{-1} f(x_{k-1}, y_{k-1}) =$

   $\begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} - \begin{bmatrix} \frac{1}{10x_{k-1}} & \frac{1}{10x_{k-1}} \\ \frac{2}{5y_{k-1}} & -\frac{1}{10y_{k-1}} \end{bmatrix} \begin{bmatrix} x_{k-1}^2 + y_{k-1}^2 - 4 \\ 4x_{k-1}^2 - y_{k-1}^2 - 4 \end{bmatrix}$

   $\begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} - \begin{bmatrix} \frac{5x_{k-1}^2-8}{10x_{k-1}} \\ \frac{5y_{k-1}^2-12}{10y_{k-1}} \end{bmatrix} = \begin{bmatrix} x_{k-1} - \frac{5x_{k-1}^2-8}{10x_{k-1}} \\ y_{k-1} - \frac{5y_{k-1}^2-12}{10y_{k-1}} \end{bmatrix}$

   Beginning with $x_0 = y_0 = 1$, $x_1 = 1 - \frac{5.1^2-8}{10.1} = 1.3$, $y_1 = 1 - \frac{5.1^2-12}{10.1} = 1.7$

Appendix: some possibly useful formula

- Lagrange mechanics is built on the equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{q}} = \mathbf{Q}, \qquad \mathcal{L}\left(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{z}\right) = \mathcal{T} - \mathcal{V} - \mathbf{z}^\top \mathbf{C}, \qquad \mathbf{C} = 0, \qquad \langle \delta\mathbf{q}, \mathbf{Q}\rangle = \delta W, \forall\, \delta\mathbf{q} \tag{22}$$

  The kinetic and potential energy of a point mass are given by:

$$\mathcal{T} = \frac{1}{2}m\dot{\mathbf{p}}^\top\dot{\mathbf{p}}, \qquad \mathcal{V} = mg\mathbf{p}_3 \tag{23}$$

  respectively, where $\mathbf{p} \in \mathbb{R}^3$ is the position of the mass in a cartesian reference frame having the third coordinate as the vertical axis pointing up. The generalized forces are identical to the external forces applied to a point mass if the position of that point is expressed in cartesian coordinates in the generalized coordinates $\mathbf{q}$.

- In the case $\mathcal{T} = \frac{1}{2}m\dot{\mathbf{q}}^\top W\dot{\mathbf{q}}$ with W constant $\mathcal{V} = \mathcal{V}\left(\mathbf{q}\right)$ and $\mathbf{C} = \mathbf{C}\left(\mathbf{q}\right)$, the Lagrange equations simplify to the dynamics in the semi-explicit index-3 DAE form:

$$\dot{\mathbf{p}} = \mathbf{v} \tag{24a}$$

$$W\dot{\mathbf{v}} + \frac{\partial \mathbf{C}}{\partial \mathbf{q}}^\top \mathbf{z} = \mathbf{Q} - \frac{\partial \mathcal{V}}{\partial \mathbf{q}}^\top \tag{24b}$$

$$0 = \mathbf{C}\left(\mathbf{q}\right) \tag{24c}$$

- The Implicit Function Theorem (IFT) guarantees that a nonlinear set of equations

$$\mathbf{r}\left(\mathbf{y}, \mathbf{z}\right) = 0 \tag{25}$$

  "can be solved" in terms of $\mathbf{z}$ for a given $\mathbf{y}$ iff the Jacobian $\frac{\partial \mathbf{r}(\mathbf{y},\mathbf{z})}{\partial \mathbf{z}}$ is full rank at the solution. More specifically, it guarantees that there is a function $\phi\left(\mathbf{y}\right)$ such that

$$\mathbf{r}\left(\mathbf{y}, \phi\left(\mathbf{y}\right)\right) = 0 \tag{26}$$

  holds in the neighborhood of the point $\mathbf{y}$ where the Jacobian is evaluated. Furthermore, the IFT specifies that:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = -\frac{\partial \mathbf{r}}{\partial \mathbf{z}}^{-1}\frac{\partial \mathbf{r}}{\partial \mathbf{y}} \tag{27}$$

- For solving a problem $\mathbf{r}\left(\mathbf{x}\right) = 0$, Newton iterates:

$$\mathbf{x} \leftarrow \mathbf{x} - \alpha\frac{\partial \mathbf{r}}{\partial \mathbf{x}}^{-1}\mathbf{r} \tag{28}$$

  until $\mathbf{r}\left(\mathbf{x}\right) \approx 0$ where $\alpha \in [0, 1]$

- Runge-Kutta methods are described by:

$$\begin{array}{c|ccc} c_1 & a_{11} & \dots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \dots & a_{ss} \\ \hline & b_1 & \dots & b_s \end{array}$$

$$\mathbf{K}_j = \mathbf{f}\left(\mathbf{x}_k + \Delta t\sum_{i=1}^{s} a_{ji}\mathbf{K}_i,\, \mathbf{u}\left(t_k + c_j\Delta t\right)\right), \quad j = 1, \dots, s \tag{29a}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta t\sum_{i=1}^{s} b_i\mathbf{K}_i \tag{29b}$$

- For ERK methods, the relationship between the (minimum) number of stages s to the order o is given by:

s	1	2	3	4	6	7	9	11	...
o	1	2	3	4	5	6	7	8	...

Table 1: Stage to order of ERK methods

- Collocation methods use:

$$\dot{\mathbf{x}}(t_k + \Delta t \cdot \tau) \approx \dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau) = \sum_{i=1}^{s} \mathbf{K}_i \ell_i(\tau), \quad \tau \in [0,\, 1] \tag{30}$$

$$\mathbf{x}(t_k + \Delta t \cdot \tau) \approx \hat{\mathbf{x}}(t_k + \Delta t \cdot \tau) = \mathbf{x}_k + \Delta t \sum_{i=1}^{s} \mathbf{K}_i L_i(\tau) \tag{31}$$

where the Lagrange polynomials are given by:

$$\ell_i(\tau) = \prod_{j=1, j\neq i}^{s} \frac{\tau - \tau_j}{\tau_i - \tau_j}, \quad \text{and} \quad L_i(\tau) = \int_0^{\tau} \ell_i(\xi)\mathrm{d}\xi \tag{32}$$

The Lagrange polynomials satisfy the conditions of

$$\text{Orthogonality:} \quad \int_0^1 \ell_i(\tau)\ell_j(\tau)\,\mathrm{d}\tau = 0 \quad \text{for} \quad i \neq j \tag{33a}$$

$$\text{Punctuality:} \quad \ell_i(\tau_j) = \begin{cases} 1 & \text{if} \quad j = i \\ 0 & \text{if} \quad j \neq i \end{cases} \tag{33b}$$

and enforce the collocation equations (for $j = 1, \ldots, s$):

$$\dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau_j) = \mathbf{f}\left(\hat{\mathbf{x}}(t_k + \Delta t \cdot \tau_j),\, \mathbf{u}\left(t_k + \Delta t \cdot \tau_j\right)\right), \qquad \text{in the explicit ODE case} \tag{34a}$$

$$\mathbf{F}\left(\dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau_j), \hat{\mathbf{x}}(t_k + \Delta t \cdot \tau_j),\, \mathbf{u}\left(t_k + \Delta t \cdot \tau_j\right)\right) = 0, \qquad \text{in the implicit ODE case} \tag{34b}$$

$$\mathbf{F}\left(\dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau_j), \hat{\mathbf{z}}_j, \hat{\mathbf{x}}(t_k + \Delta t \cdot \tau_j),\, \mathbf{u}\left(t_k + \Delta t \cdot \tau_j\right)\right) = 0, \qquad \text{in the fully-implicit DAE case} \tag{34c}$$

- Gauss-Legendre collocation methods select the set of points $\tau_{1,\ldots,s}$ as the zeros of the (shifted) Legrendre polynomial:

$$P_s\left(\tau\right) = \frac{1}{s!}\frac{\mathrm{d}^s}{\mathrm{d}\tau^s}\left[\left(\tau^2 - \tau\right)^s\right] \tag{35}$$

They achieve the order $\|\mathbf{x}_N - \mathbf{x}\left(t_{\mathrm{f}}\right)\| = \mathcal{O}\left(\Delta t^{2s}\right)$.

- Maximum-likelihood estimation is based on

$$\max_{\boldsymbol{\theta}} \quad \mathbb{P}\left[e_k = y_k - \hat{y}_k \quad \text{for} \quad k = 1, \ldots, N \mid \boldsymbol{\theta}\right] \tag{36}$$

If the noise sequence is uncorrelated, then

$$\mathbb{P}\left[e_k = y_k - \hat{y}_k \quad \text{for} \quad k = 0, \ldots, N \mid \boldsymbol{\theta}\right] = \prod_{k=1}^{N} \mathbb{P}\left[e_k = y_k - \hat{y}_k \mid \boldsymbol{\theta}\right] \tag{37}$$

- The solution of a linear least-squares problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\left\|A\boldsymbol{\theta} - \mathbf{y}\right\|^2_{\Sigma_e^{-1}} \tag{38}$$

reads as:

$$\hat{\boldsymbol{\theta}} = \left(A^\top \Sigma_e^{-1} A\right)^{-1} A^\top \Sigma_e^{-1} \mathbf{y} \tag{39}$$

and the covariance of the parameter estimation based is given by the formula:

$$\Sigma_{\hat{\boldsymbol{\theta}}} = \left(A^\top \Sigma_e^{-1} A\right)^{-1} \tag{40}$$

- In system identification, given the a plant $G(z)$ and a noise $H(z)$ model description, the one-step-ahead predictor $\hat{y}(k|k-1)$ can be retrieved with

$$H(z)\hat{y}(z) = G(z)u(z) + (H(z) - 1)y(z) \tag{41}$$

- The Gauss-Newton approximation in an optimization problem

$$\min_{\mathbf{x}} \quad J(\mathbf{x}) = \frac{1}{2} \left\|\mathbf{R}(\mathbf{x})\right\|^2 \tag{42}$$

uses the approximation:

$$\frac{\partial^2 J}{\partial \mathbf{x}^2} \approx \frac{\partial R}{\partial \mathbf{x}}^\top \frac{\partial R}{\partial \mathbf{x}} \tag{43}$$

- The solution to an LTI system $\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}$ is given by:

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) + \int_0^t e^{A(t-\tau)} B\mathbf{u}(\tau)\mathrm{d}\tau \tag{44}$$

and the transformation state-space to transfer function is given by:

$$G(s) = C\left(sI - A\right)^{-1} B + D \tag{45}$$

- $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $det(A) = ad - bc$
- $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, $det(A) = a.det(\begin{bmatrix} e & f \\ h & i \end{bmatrix}) - b.det(\begin{bmatrix} d & f \\ g & i \end{bmatrix}) + c.det(\begin{bmatrix} d & e \\ g & h \end{bmatrix})$
- $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $det(A) = ad - bc$, $A^{-1} = \frac{1}{det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
- $\alpha = \mathbf{x}^T A\mathbf{x}$, where A is a symmetric matrix and $\mathbf{x}$ is $n \times 1$, A is $n \times n$, and A does not depend on $\mathbf{x}$, then, $\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}^T A$.