Modelling and Simulation ESS101
27 October 2021 (Final exam)

This exam contains 11 pages (including this cover page) and 4 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 a priori be given at 24 points, and the top grade at 36 points. These limits may be lowered depending on the outcome of the exam.

Problem	Points	Score
1	11	
2	10	
3	11	
4	8	
Total:	40	

Best of luck to all !!

1. Consider two masses of mass m (the black balls in Fig. 1) linked by a massless rigid rod of length L. The masses are gliding without friction on a surface of equation:

$$z = \frac{1}{2}x^2 + \frac{1}{2}y^2 \tag{1}$$

(a) (4 points) Write the Lagrange function describing the system, assuming gravity is the only external force.

(b) (4 points) Derive the equations of motion

(c) (3 points) What are the consistency conditions associated to the equations?

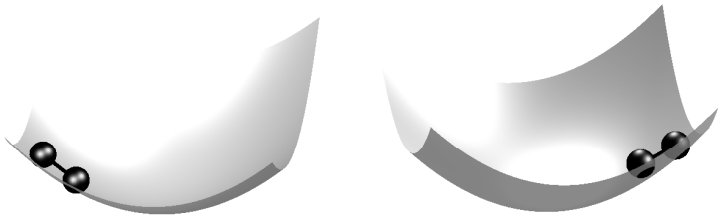

Figure 1: Illustration of the rollercoaster

Solution:

(a) The Lagrange function comprises the kinetic (T) and potential energy (V) of the balls, and the constraint (1). Let us describe the position of the balls via

$$\mathbf{q}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}, \qquad \mathbf{q}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \tag{2}$$

We then have (to avoid confusion with the coordinate z, $\boldsymbol{\lambda}$ is used instead of $\mathbf{z}$ for the Lagrange multipliers):

$$T = \frac{1}{2}m\dot{\mathbf{q}}_1^\top \dot{\mathbf{q}}_1 + \frac{1}{2}m\dot{\mathbf{q}}_2^\top \dot{\mathbf{q}}_2, \qquad V = mg\,(z_1 + z_2) \tag{3a}$$

$$\mathcal{L} = T - V + \boldsymbol{\lambda}^\top \begin{bmatrix} \frac{1}{2}x_1^2 + \frac{1}{2}y_1^2 - z_1 \\ \frac{1}{2}x_2^2 + \frac{1}{2}y_2^2 - z_2 \\ \frac{1}{2}\left(\|\mathbf{q}_1 - \mathbf{q}_2\|^2 - L^2\right) \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \tag{3b}$$

(b) The equations of motion are given by a simple application of the Lagrange equations, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{q}} = 0, \tag{4}$$

where the generalized coordinates are $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2) = (x_1, y_1, z_1, x_2, y_2, z_2)$.

We can therefore compute:

$$\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}}^\top = m \begin{bmatrix} \ddot{\mathbf{q}}_1 \\ \ddot{\mathbf{q}}_2 \end{bmatrix}, \qquad \frac{\partial \mathcal{L}}{\partial \mathbf{q}}^\top = \begin{bmatrix} \lambda_1 x_1 \\ \lambda_1 y_1 \\ -\lambda_1 - gm \\ \lambda_2 x_2 \\ \lambda_2 y_2 \\ -\lambda_2 - gm \end{bmatrix} + \lambda_3 \begin{bmatrix} \mathbf{q}_1 - \mathbf{q}_2 \\ \mathbf{q}_2 - \mathbf{q}_1 \end{bmatrix} \tag{5a}$$

such that the equations of motion read as:

$$m \begin{bmatrix} \ddot{\mathbf{q}}_1 \\ \ddot{\mathbf{q}}_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 \\ \lambda_1 y_1 \\ -\lambda_1 - gm \\ \lambda_2 x_2 \\ \lambda_2 y_2 \\ -\lambda_2 - gm \end{bmatrix} + \lambda_3 \begin{bmatrix} \mathbf{q}_1 - \mathbf{q}_2 \\ \mathbf{q}_2 - \mathbf{q}_1 \end{bmatrix} \tag{6}$$

(c) Let us define:

$$\mathbf{c} = \begin{bmatrix} \frac{1}{2}x_1^2 + \frac{1}{2}y_1^2 - z_1 \\ \frac{1}{2}x_2^2 + \frac{1}{2}y_2^2 - z_2 \\ \frac{1}{2}\left(\|\mathbf{q}_1 - \mathbf{q}_2\|^2 - L^2\right) \end{bmatrix} \tag{7}$$

The consistency conditions require that the initial conditions of the system must satisfy $\mathbf{c} = 0$ and $\dot{\mathbf{c}} = 0$. The latter equation requires that:

$$\begin{bmatrix} x_1\dot{x}_1 + y_1\dot{y}_1 - \dot{z}_1 \\ x_2\dot{x}_2 + y_2\dot{y}_2 - \dot{z}_2 \\ (\mathbf{q}_1 - \mathbf{q}_2)^\top (\dot{\mathbf{q}}_1 - \dot{\mathbf{q}}_2) \end{bmatrix} = 0 \tag{8}$$

holds on the initial conditions.

2. (a) (5 points) Consider the differential equation:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \dot{\mathbf{x}} = \mathbf{x} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u \tag{9}$$

1. Is (9) an implicit ODE or a DAE? Justify.
2. If it is an implicit ODE, what is its solution for $u = 0$? If it is a DAE, what is its differential index?

(b) (5 points) Perform an index reduction of the DAE:

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + \mathbf{b}z \\ 0 &= \frac{1}{2}\left(\mathbf{x}^\top \mathbf{x} - L^2\right) \end{aligned}$$

where L may be time-varying. What are the consistency conditions? What condition is needed on $\mathbf{b}$ and L for the DAE to be well-posed?

Solution:

(a) 1. Since matrix E is rank defficient, (9) is a DAE.

2. We observe that (9) reads as:

$$\dot{\mathbf{x}}_2 = \mathbf{x}_1 + u \tag{10a}$$
$$\dot{\mathbf{x}}_3 = \mathbf{x}_2 \tag{10b}$$
$$0 = \mathbf{x}_3 \tag{10c}$$

and is a semi-explicit DAE where $\mathbf{x}_1$ is the algebraic variable. It cannot be computed readily from (10) (equivalently the Jacobian of the algebraic constraint (10c) with respect to $\mathbf{x}_1$ is zero, i.e. rank deficient), hence we have a "high-index DAE" (index more than 1). In order to assess the index, we perform a $\frac{\mathrm{d}}{\mathrm{d}t}$ on the algebraic constraint (10c) (the other equations are explicit ODEs, we leave them alone). We obtain

$$\dot{\mathbf{x}}_3 = 0 \tag{11}$$

and use (10b) to get the new algebraic constraint $\mathbf{x}_2 = 0$. We perform a new $\frac{\mathrm{d}}{\mathrm{d}t}$, to obtain

$$\dot{\mathbf{x}}_2 = 0 \tag{12}$$

and use (10a) to obtain the new algebraic constraint:

$$\mathbf{x}_1 + u = 0 \tag{13}$$

We need yet another $\frac{\mathrm{d}}{\mathrm{d}t}$ to obtain:

$$\dot{\mathbf{x}}_1 + \dot{u} = 0 \tag{14}$$

The ODE corresponding to (9) is therefore:

$$\dot{\mathbf{x}}_2 = \mathbf{x}_1 + u \tag{15a}$$
$$\dot{\mathbf{x}}_3 = \mathbf{x}_2 \tag{15b}$$
$$\dot{\mathbf{x}}_1 = -\dot{u} \tag{15c}$$
$$\tag{15d}$$

Since we have performed 3 time differentiation to get there, our DAE is of index 3.

(b) We perform the index reduction by taking the time derivative of the algebraic constraint $g = \frac{1}{2}\left(\mathbf{x}^\top\mathbf{x} - L^2\right)$:

$$\dot{g} = \mathbf{x}^\top\dot{\mathbf{x}} - L(t)\dot{L}(t) = \mathbf{x}^\top A\mathbf{x} + \mathbf{x}^\top\mathbf{b}z - L(t)\dot{L}(t) = 0 \tag{16}$$

which is solvable for z as long as $\mathbf{x}^\top\mathbf{b} \neq 0$. We then have the index-1 DAE

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{b}z \tag{17a}$$
$$0 = \mathbf{x}^\top A\mathbf{x} + \mathbf{x}^\top\mathbf{b}z - L(t)\dot{L}(t) \tag{17b}$$

with the consistency condition $g(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}^\top\mathbf{x} - L^2\right) = 0$. The consistency condition is simply $\frac{1}{2}\left(\mathbf{x}(0)^\top\mathbf{x}(0) - L(0)^2\right) = 0$. For $L = 0$, constraint entails that $\mathbf{x} = 0$, where $\mathbf{x}^\top B \neq 0$ fails. For the DAE to be well-posed we need $\mathbf{x}^\top\mathbf{b} \neq 0$, i.e. the trajectory of $\mathbf{x}$ cannot end-up at a point where the vector $\mathbf{x}$ is orthogonal to the vector $\mathbf{b}$.

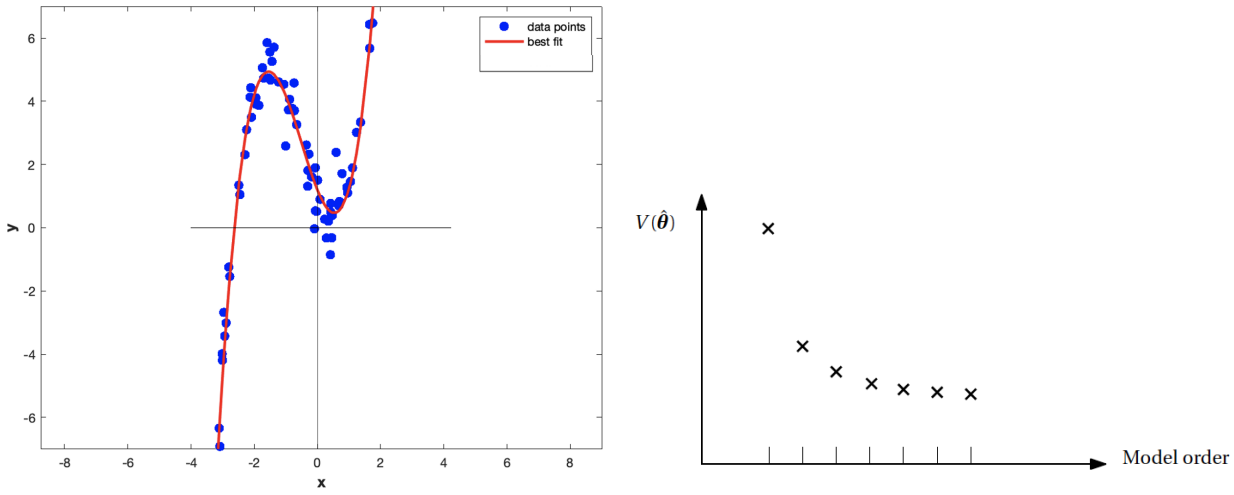

Figure 2: Linear regression based on least squares and Loss function for increasing model orders.

3. (a) (7 points) Consider a two-dimensional dataset $\{x(i), y(i)\}, i = 1, ...N$ illustrated in Fig. 2 on the left.

   1. Using a least squares approach capturing the relationship between x and y,

   $$\hat{\mathbf{y}} = a + bx + cx^2 + dx^3 = \boldsymbol{\theta}^\top \boldsymbol{\varphi},$$

   where $\boldsymbol{\theta}$ is the parameter vector and $\boldsymbol{\varphi}$ is the *regression vector* (holding the *regressors* $1, x, ...$), write out all the components in the R_N and f_N matrices in the least squares estimate defined as,

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

   which would correspond to the fitted red curve in Fig. 2 on the left.
   2. Write the general form of these R_N and f_N matrices that correspond to a predictor $\hat{\mathbf{y}}$ based on polynomial with degree k.
   3. Discuss what overfitting to training data means and discuss an approach to avoid overfitting based on the relation between the loss function and model order as in Fig. 2 on the right, commenting on how to make use of this relation.

   (b) (4 points) Consider the model structure

   $$y(t) + \alpha y(t-1) = u(t-1) + e(t) + \gamma e(t-1), \tag{18}$$

   where $e(\cdot)$ is a sequence of i.i.d. random variables with zero mean, and the coefficient for u is known to be 1.

   1. What kind of model structure is it?
   2. Compute the one-step ahead predictor $\hat{y}(t|t-1)$ for the model (18).
   3. We want to fit the above model to data by minimizing the quadratic criterion

   $$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}(y(t) - \hat{y}(t|t-1))^2 \tag{19}$$

   with $\theta^T = \begin{bmatrix}\alpha & \gamma\end{bmatrix}$. How can the minimizing $\boldsymbol{\theta}$ be found? Motivate your solution.

4. Suggest how the model structure (18) could be modified to simplify computations considerably. Motivate!

Solution:

(a) 1.

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^{\top}(i) = \frac{1}{N}\begin{bmatrix} N & \sum_{i=1}^{N}x(i) & \sum_{i=1}^{N}x^2(i) & \sum_{i=1}^{N}x^3(i) \\ \sum_{i=1}^{N}x(i) & \sum_{i=1}^{N}x^2(i) & \sum_{i=1}^{N}x^3(i) & \sum_{i=1}^{N}x^4(i) \\ \sum_{i=1}^{N}x^2(i) & \sum_{i=1}^{N}x^3(i) & \sum_{i=1}^{N}x^4(i) & \sum_{i=1}^{N}x^5(i) \\ \sum_{i=1}^{N}x^3(i) & \sum_{i=1}^{N}x^4(i) & \sum_{i=1}^{N}x^5(i) & \sum_{i=1}^{N}x^6(i) \end{bmatrix} \tag{20}$$

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)y(i) = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N}y(i) \\ \sum_{i=1}^{N}x(i)y(i) \\ \sum_{i=1}^{N}x^2(i)y(i) \\ \sum_{i=1}^{N}x^3(i)y(i) \end{bmatrix} \tag{21}$$

2.

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^{\top}(i) = \frac{1}{N}\begin{bmatrix} N & \sum_{i=1}^{N}x(i) & ... & \sum_{i=1}^{N}x^k(i) \\ \sum_{i=1}^{N}x(i) & \sum_{i=1}^{N}x^2(i) & ... & \sum_{i=1}^{N}x^{k+1}(i) \\ . & . & . & \\ . & . & . & \\ . & . & . & \\ \sum_{i=1}^{N}x^k(i) & \sum_{i=1}^{N}x^{k+1}(i) & ... & \sum_{i=1}^{N}x^{2*k}(i) \end{bmatrix} \tag{22}$$

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(i)y(i) = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N}y(i) \\ \sum_{i=1}^{N}x(i)y(i) \\ . \\ . \\ . \\ \sum_{i=1}^{N}x^k(i)y(i) \end{bmatrix} \tag{23}$$

3. Overfitting occurs when the model parameters are too fine tuned wrt training data and the model complexity is too high, this means the model predictions do not perform well when there is unseen test data. We can choose model order when the loss does not drop significantly any more - corresponding to the knee in the figure.

(b) Using the backward shift operator q^{-1} (alternatively, z^{-1} can be used), the model can equivalently be written as

$$(1+\alpha q^{-1})y(t) = q^{-1}u(t) + (1+\gamma q^{-1})e(t). \tag{24}$$

1. ARMAX.
2. Rewrite the model as

$$(1+\gamma q^{-1})y(t) = (\gamma-\alpha)q^{-1}y(t) + q^{-1}u(t) + (1+\gamma q^{-1})e(t), \tag{25}$$

or, equivalently,

$$y(t) = \frac{1}{1+\gamma q^{-1}}\left[(\gamma-\alpha)q^{-1}y(t) + q^{-1}u(t)\right] + e(t) = \hat{y}(t|t-1) + e(t), \tag{26}$$

where the latter equality follows from the fact that $e(t)$ is the only part of $y(t)$ that cannot

be predicted. Hence, the predictor is given by the difference equation

$$(1 + \gamma q^{-1})\hat{y}(t|t-1) = (\gamma - \alpha)y(t-1) + u(t-1) \tag{27}$$

3. Since $\hat{y}$ depends nonlinearly on $\boldsymbol{\theta}$, the minimization is approached by iteratively searching for a solution to the equation
$$\nabla_{\boldsymbol{\theta}} V_N(\boldsymbol{\theta}) = 0$$

4. By choosing $\gamma = 0$, the model structure becomes an ARX model, and the predictor becomes linear in $\boldsymbol{\theta}$. The minimizing θ can then be computed explicitly.

4. (a) (2 points) Specify what information the Butcher tableau readily provides on the resulting RK scheme, and what information is not obviously available (answers can be short but must be specific). point is given for a short answer, an extra point for a more detailed discussion.

   (b) (2 points) Why are high-order explicit RK methods often not the optimal choice?

   (c) (4 points) Consider the following Runge-Kutta equations for integration of an ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$:

$$\mathbf{K}_1 = \mathbf{f}\left(\mathbf{x}_k, \mathbf{u}(t_k)\right) \tag{28a}$$

$$\mathbf{K}_2 = \mathbf{f}\big(\mathbf{x}_k + \frac{\Delta t}{2}(\mathbf{K}_1 + \mathbf{K}_2), \mathbf{u}(t_k + \Delta t)\big) \tag{28b}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\Delta t}{2}(\mathbf{K}_1 + \mathbf{K}_2) \tag{28c}$$

   1. Determine, if possible, the number of stages and the order of the scheme, and whether it is an explicit or implicit RK scheme.
   2. What is the Butcher array describing the scheme?
   3. Determine the stability function.
   4. Is the scheme A-stable?

Solution:

(a) The Butcher tableau specifies: the number of stages of the RK method, whether the method is implicit or explicit and enough information to code the RK scheme. It does not (readily) provide the order of the integration method, as the order depends on the specific entries used in the tableau. To assess the order from the tableau, one needs to perform (possibly) involved computations.

(b) Up to order $o = 4$, ERK methods require $s = o$ stages, hence $s = o$ evaluations of the model equations. Each extra function evaluation readily delivers an extra order of accuracy, and allows for reducing the total number of function evaluation required. This trend is broken for $o > 4$. At higher orders, the required number of stages (and hence the number of function evaluations) progresses faster than o. Then the overall computational cost of obtaining a given accuracy tends to not improve (or even increase) for higher orders.

(c)
1. The RK scheme is implicit and has 2 stages, but the order cannot be determined in a straightforward way.
2. The Butcher array is given by

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

3. Denoting the Butcher array as

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

the stability function is given by $R(\mu) = 1 + \mu b^T (I - \mu A)^{-1}\mathbf{1}$, where $\mu = \lambda \Delta t$ and $\mathbf{1}$ is a column vector with all entries equal to 1. Thus:

$$R(\mu) = 1 + \mu \begin{bmatrix} 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\mu/2 & 1 - \mu/2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1 + \mu/2}{1 - \mu/2}$$

4. Since $|1 + \mu/2| \leq |1 - \mu/2|$ for all μ in the left half-plane, $|R(\mu)| \leq 1$ for the same μ, i.e. the scheme is A-stable.

Appendix: some possibly useful formula

- Lagrange mechanics is built on the equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{q}} = \mathbf{Q}, \qquad \mathcal{L}\left(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{z}\right) = \mathcal{T} - \mathcal{V} - \mathbf{z}^\top \mathbf{C}, \qquad \mathbf{C} = 0, \qquad \langle \delta\mathbf{q},\, \mathbf{Q}\rangle = \delta W, \forall\, \delta\mathbf{q} \tag{29}$$

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

$$\mathcal{T} = \frac{1}{2} m \dot{\mathbf{p}}^\top \dot{\mathbf{p}}, \qquad \mathcal{V} = mg\mathbf{p}_3 \tag{30}$$

  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:

$$\dot{\mathbf{p}} = \mathbf{v} \tag{31a}$$

$$W\dot{\mathbf{v}} + \frac{\partial \mathbf{C}}{\partial \mathbf{q}}^\top \mathbf{z} = \mathbf{Q} - \frac{\partial \mathcal{V}}{\partial \mathbf{q}}^\top \tag{31b}$$

$$0 = \mathbf{C}\left(\mathbf{q}\right) \tag{31c}$$

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

$$\mathbf{r}\left(\mathbf{y}, \mathbf{z}\right) = 0 \tag{32}$$

  "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

$$\mathbf{r}\left(\mathbf{y}, \phi\left(\mathbf{y}\right)\right) = 0 \tag{33}$$

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

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = -\frac{\partial \mathbf{r}}{\partial \mathbf{z}}^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{y}} \tag{34}$$

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

$$\mathbf{x} \leftarrow \mathbf{x} - \alpha \frac{\partial \mathbf{r}}{\partial \mathbf{x}}^{-1} \mathbf{r} \tag{35}$$

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

- Runge-Kutta methods are described by:

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

$$\mathbf{K}_j = \mathbf{f}\left(\mathbf{x}_k + \Delta t \sum_{i=1}^{s} a_{ji}\mathbf{K}_i,\, \mathbf{u}\left(t_k + c_j \Delta t\right)\right), \quad j = 1, \dots, s \tag{36a}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta t \sum_{i=1}^{s} b_i \mathbf{K}_i \tag{36b}$$

- 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:

$$\dot{\mathbf{x}}(t_k + \Delta t \cdot \tau) \approx \dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau) = \sum_{i=1}^{s} \mathbf{K}_i \ell_i(\tau), \quad \tau \in [0,\, 1] \tag{37}$$

$$\mathbf{x}(t_k + \Delta t \cdot \tau) \approx \hat{\mathbf{x}}(t_k + \Delta t \cdot \tau) = \mathbf{x}_k + \Delta t \sum_{i=1}^{s} \mathbf{K}_i L_i(\tau) \tag{38}$$

where the Lagrange polynomials are given by:

$$\ell_i(\tau) = \prod_{j=1, j\neq i}^{s} \frac{\tau - \tau_j}{\tau_i - \tau_j}, \quad \text{and} \quad L_i(\tau) = \int_0^{\tau} \ell_i(\xi) \mathrm{d}\xi \tag{39}$$

The Lagrange polynomials satisfy the conditions of

$$\text{Orthogonality:} \quad \int_0^1 \ell_i(\tau)\ell_j(\tau)\,\mathrm{d}\tau = 0 \quad \text{for} \quad i \neq j \tag{40a}$$

$$\text{Punctuality:} \quad \ell_i(\tau_j) = \begin{cases} 1 & \text{if} \quad j = i \\ 0 & \text{if} \quad j \neq i \end{cases} \tag{40b}$$

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

$$\dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau_j) = \mathbf{f}\left(\hat{\mathbf{x}}(t_k + \Delta t \cdot \tau_j),\, \mathbf{u}\left(t_k + \Delta t \cdot \tau_j\right)\right), \qquad \text{in the explicit ODE case} \tag{41a}$$

$$\mathbf{F}\left(\dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau_j), \hat{\mathbf{x}}(t_k + \Delta t \cdot \tau_j),\, \mathbf{u}\left(t_k + \Delta t \cdot \tau_j\right)\right) = 0, \qquad \text{in the implicit ODE case} \tag{41b}$$

$$\mathbf{F}\left(\dot{\hat{\mathbf{x}}}(t_k + \Delta t \cdot \tau_j), \hat{\mathbf{z}}_j, \hat{\mathbf{x}}(t_k + \Delta t \cdot \tau_j),\, \mathbf{u}\left(t_k + \Delta t \cdot \tau_j\right)\right) = 0, \qquad \text{in the fully-implicit DAE case} \tag{41c}$$

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

$$P_s\left(\tau\right) = \frac{1}{s!}\frac{\mathrm{d}^s}{\mathrm{d}\tau^s}\left[\left(\tau^2 - \tau\right)^s\right] \tag{42}$$

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

$$\max_{\boldsymbol{\theta}} \quad \mathbb{P}\left[e_k = y_k - \hat{y}_k \quad \text{for} \quad k = 1, \ldots, N \mid \boldsymbol{\theta}\right] \tag{43}$$

If the noise sequence is uncorrelated, then

$$\mathbb{P}\left[e_k = y_k - \hat{y}_k \quad \text{for} \quad k = 0, \ldots, N \mid \boldsymbol{\theta}\right] = \prod_{k=1}^{N} \mathbb{P}\left[e_k = y_k - \hat{y}_k \,|\, \boldsymbol{\theta}\,\right] \tag{44}$$

- The solution of a linear least-squares problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\left\|A\boldsymbol{\theta} - \mathbf{y}\right\|_{\Sigma_e^{-1}}^2 \tag{45}$$

reads as:

$$\hat{\boldsymbol{\theta}} = \left(A^\top \Sigma_e^{-1} A\right)^{-1} A^\top \Sigma_e^{-1} \mathbf{y} \tag{46}$$

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

$$\Sigma_{\hat{\boldsymbol{\theta}}} = \left(A^\top \Sigma_e^{-1} A\right)^{-1} \tag{47}$$

- 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

$$H(z)\hat{y}(z) = G(z)u(z) + (H(z) - 1)y(z) \tag{48}$$

- The Gauss-Newton approximation in an optimization problem

$$\min_{\mathbf{x}} \quad J(\mathbf{x}) = \frac{1}{2} \|\mathbf{R}(\mathbf{x})\|^2 \tag{49}$$

uses the approximation:

$$\frac{\partial^2 J}{\partial \mathbf{x}^2} \approx \frac{\partial R}{\partial \mathbf{x}}^\top \frac{\partial R}{\partial \mathbf{x}} \tag{50}$$

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

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) + \int_0^t e^{A(t-\tau)} B\mathbf{u}(\tau) \mathrm{d}\tau \tag{51}$$

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

$$G(s) = C\left(sI - A\right)^{-1} B + D \tag{52}$$