To do so, we employ the *Cayley-Hamilton theorem* from linear algebra. Because $\boldsymbol{A}$ is $N \times N$, its characteristic equation has at most N terms, and therefore any power of $\boldsymbol{A}$ greater than or equal to N can be rewritten as a linear combination of $\mathbf{1}, \boldsymbol{A}, \ldots, \boldsymbol{A}^{(N-1)}$. By extension, for any $k \geq N$, we can write

Cayley-Hamilton theorem: Every square matrix, $\boldsymbol{A}$, over the real field, satisfies its own characteristic equation, $\det(\lambda\mathbf{1} - \boldsymbol{A}) = 0$.

$$\begin{aligned}&\left(\boldsymbol{A}^T\right)^{(k-1)} \boldsymbol{C}^T \\ &= a_0 \mathbf{1}^T \boldsymbol{C}^T + a_1 \boldsymbol{A}^T \boldsymbol{C}^T + a_2 \boldsymbol{A}^T \boldsymbol{A}^T \boldsymbol{C}^T + \cdots + a_{N-1} \left(\boldsymbol{A}^T\right)^{(N-1)} \boldsymbol{C}^T\end{aligned} \tag{3.44}$$

for some set of scalars, $a_0, a_1, \ldots, a_{N-1}$, not all zero. Since row-rank and column-rank are the same for any matrix, we can conclude that

$$\begin{aligned}\operatorname{rank} &\begin{bmatrix} \boldsymbol{C}^T & \boldsymbol{A}^T \boldsymbol{C}^T & \boldsymbol{A}^T \boldsymbol{A}^T \boldsymbol{C}^T & \cdots & \left(\boldsymbol{A}^T\right)^K \boldsymbol{C}^T \end{bmatrix} \\ &= \operatorname{rank} \begin{bmatrix} \boldsymbol{C}^T & \boldsymbol{A}^T \boldsymbol{C}^T & \cdots & \left(\boldsymbol{A}^T\right)^{(N-1)} \boldsymbol{C}^T \end{bmatrix}.\end{aligned} \tag{3.45}$$

Defining the *observability matrix*, $\mathcal{O}$, as

$$\mathcal{O} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{(N-1)} \end{bmatrix}, \tag{3.46}$$

our rank condition is

$$\operatorname{rank} \mathcal{O} = N. \tag{3.47}$$

Readers familiar with linear control theory will recognize this as precisely the test for *observability* (Kalman, 1960a). Thus, we can see the direct connection between observability and invertibility of $\mathbf{H}^T \mathbf{W}^{-1} \mathbf{H}$. The overall conditions for existence and uniqueness of a solution to (3.34) are

A system is *observable* if the initial state can be uniquely inferred based on measurements gathered in a finite amount of time.

$$\mathbf{Q}_k > 0, \quad \mathbf{R}_k > 0, \quad \operatorname{rank} \mathcal{O} = N, \tag{3.48}$$

where > 0 means a matrix is positive-definite (and hence invertible). Again, these are sufficient but not necessary conditions.

Interpretation

We can return to the mass-spring analogy to better understand the observability issue. Figure 3.2 shows a few examples. With the initial state and all the inputs (top example), the system is always observable since it is impossible to move any group of carts left or right without altering the length of at least one spring. This means there is a unique minimum-energy state. The same is true for the middle example, even though there is no knowledge of the initial state. The bottom example