Separable Models, Introduction to Dynamic Systems

Roland Hostettler

October 3, 2018

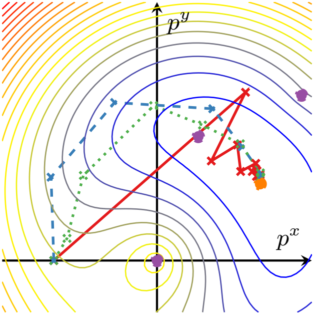
Recap

- Gauss–Newton algorithm:

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \hat{\boldsymbol{\theta}}^{(i)} + \gamma(\boldsymbol{G}_{\boldsymbol{\theta}}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}})^{-1}\boldsymbol{G}_{\boldsymbol{\theta}}^{\mathsf{T}}\boldsymbol{R}^{-1}(\boldsymbol{y} - g(\hat{\boldsymbol{\theta}}^{(i)}))$$

- Levenberg–Marquardt algorithm:

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \hat{\boldsymbol{\theta}}^{(i)} + (\boldsymbol{G}_{\boldsymbol{\theta}}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{G}_{\boldsymbol{\theta}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{G}_{\boldsymbol{\theta}}^{\mathsf{T}}\boldsymbol{R}^{-1}(\boldsymbol{y} - g(\hat{\boldsymbol{\theta}}^{(i)}))$$



Intended Learning Outcomes

After this lecture, you will be able to:

- identify and describe separable models,
- recognize the problems with dynamic systems,
- describe the idea of state-space modeling.

Example: Localization using Signal Strength

- Measurement model:

$$
\begin{aligned}
y_n &= P_{0,\text{dBm}} - 10\beta \log_{10}(|\Delta p_n|) + r_n \\
&= P_{0,\text{dBm}} - \frac{10\beta}{\log(10)} \log(|\Delta p_n|) + r_n
\end{aligned}
$$

  with $\Delta p_n = p_t - p_n$

- Parameters are $\boldsymbol{\theta} = \begin{bmatrix} P_{0,\text{dBm}} & \beta & p_t^x & p_t^y \end{bmatrix}^\mathsf{T}$
- The Jacobian is given by

$$
\boldsymbol{G}_{\boldsymbol{\theta}} = \begin{bmatrix}
1 & -10\log_{10}(|\Delta p_1|) & \frac{-10\beta}{\log(10)|\Delta p_1|^2}\Delta p_n^x & \frac{-10\beta}{\log(10)|\Delta p_1|^2}\Delta p_n^y \\
1 & -10\log_{10}(|\Delta p_2|) & \frac{-10\beta}{\log(10)|\Delta p_2|^2}\Delta p_n^x & \frac{-10\beta}{\log(10)|\Delta p_2|^2}\Delta p_n^y \\
\vdots & \vdots & \vdots & \vdots \\
1 & -10\log_{10}(|\Delta p_N|) & \frac{-10\beta}{\log(10)|\Delta p_N|^2}\Delta p_n^x & \frac{-10\beta}{\log(10)|\Delta p_N|^2}\Delta p_n^y
\end{bmatrix}
$$

Separable Models: Motivation & Definition

- Numerical optimization methods:
  - Can handle arbitrary nonlinear least squares problem
  - Are computationally expensive
  - Can be challenging to implement
- In many problems, only a few parameters enter the problem nonlinearly, the remaining ones enter linearly
- The separable model is:

$$\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2 + \boldsymbol{r},$$

  where
  - $\boldsymbol{\theta}_1$ are the nonlinear parameters
  - $\boldsymbol{\theta}_2$ the linear parameters
- Can we exploit this property?

Separable Models: Derivation (1/2)

- The separable model is:

$$\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2 + \boldsymbol{r}$$

- Weighted least squares cost function:

$$J(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2)^\mathsf{T}\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2)$$

- Weighted least squares estimator for $\boldsymbol{\theta}_2$ given $\boldsymbol{\theta}_1$:

$$\hat{\boldsymbol{\theta}}_2 = (\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{R}^{-1}\boldsymbol{y}$$

- Substitution of $\boldsymbol{\theta}_2$ in $J(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$:

$$\begin{aligned} J(\boldsymbol{\theta}_1) = (&\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{\theta}_1)(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{R}^{-1}\boldsymbol{y})^\mathsf{T}\boldsymbol{R}^{-1} \\ &\times (\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{\theta}_1)(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{R}^{-1}\boldsymbol{y}) \end{aligned}$$

Separable Models: Derivation (2/2)

- Modified cost function:

$$J(\boldsymbol{\theta}_1) = \Big(\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{\theta}_1)(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{R}^{-1}\boldsymbol{y}\Big)^\mathsf{T}\boldsymbol{R}^{-1} \\ \times \Big(\boldsymbol{y} - \boldsymbol{G}(\boldsymbol{\theta}_1)(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{R}^{-1}\boldsymbol{y}\Big)$$

- Expansion and simplification yields:

$$J(\boldsymbol{\theta}_1) = \boldsymbol{y}^\mathsf{T}\boldsymbol{y} - \boldsymbol{y}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{y}$$

- Estimating $\boldsymbol{\theta}_1$:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1 &= \underset{\boldsymbol{\theta}_1}{\operatorname{argmin}} -\boldsymbol{y}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\left(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\right)^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{y} \\ &= \underset{\boldsymbol{\theta}_1}{\operatorname{argmax}}\, \boldsymbol{y}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\left(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)\right)^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{y}\end{aligned}$$

Separable Models: Practical Considerations

- Estimating $\boldsymbol{\theta}_1$:

$$\hat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}_1}{\operatorname{argmax}}\, \boldsymbol{y}^\mathsf{T} \boldsymbol{R}^{-1} \boldsymbol{G}(\boldsymbol{\theta}_1) \left(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T} \boldsymbol{R}^{-1} \boldsymbol{G}(\boldsymbol{\theta}_1) \right)^{-1} \boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T} \boldsymbol{R}^{-1} \boldsymbol{y}$$

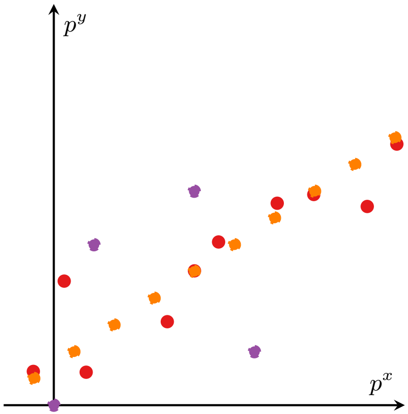
- Cost function is not in least squares form anymore (but we still solve the original least squares problem)
- Gradient descent can readily be used (but needs recalculating the gradient)
- Gauss–Newton and Levenberg–Marquardt need more work

Separable Models: Example

Localizing a Moving Target (1/4)

Assume that:

- Target moves rather than being stationary
- Sensors measure periodically, e.g. every second



Localizing a Moving Target (2/4)

- Straight line model:

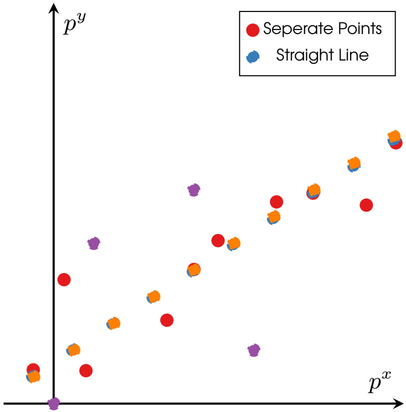
$$p_t(t) = p_t(0) + vt$$

- Measurement model

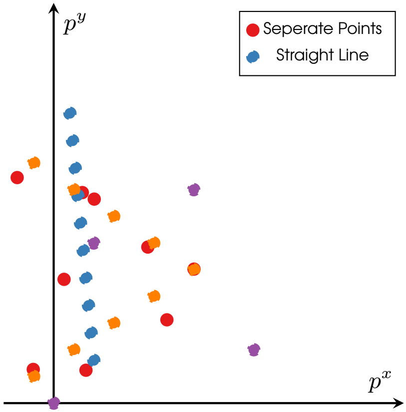
$$\begin{aligned} \boldsymbol{y}_n(t) &= |\Delta p_n(t)| + r_n(t) \\ &= |p_t(0) + vt - p_n| + r_n(t) \end{aligned}$$

- Only need to estimate 4 parameters:
  $\boldsymbol{\theta} = \begin{bmatrix} p_t^x(0) & p_t^y(0) & v^x & v^y \end{bmatrix}^\mathsf{T}$

Localizing a Moving Target (3/4)



Localizing a Moving Target (4/4)

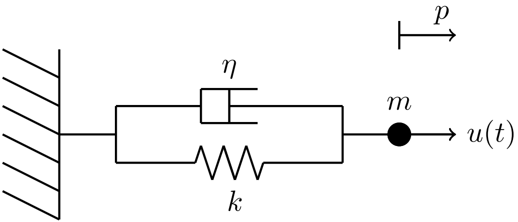


Localizing a Moving Target: Conclusions

- The static approach is not too well suited for time-varying processes
- A systematic method that relates (time-wise) related measurements is needed
- Solution: Use differential (and difference) equations to model the time-varying, i.e. dynamic, system

ODE Modeling of Dynamic Systems

- ODEs can be used to describe many dynamic systems
- Example: Spring-damper system:



- Second order ordinary differential equation:

$$ma(t) = -kp(t) - \eta v(t) + u(t)$$

- Other examples: Newtonion/Hamiltonian dynamics, kinematic models, heat and mass transfer, wave equations, ...

Example: State-Space Representation of ODEs

- Second order ordinary differential equation:

$$ma(t) = -kp(t) - \eta v(t) + u(t)$$

- Equation system representation:

$$\begin{bmatrix} v(t) \\ a(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{\eta}{m} \end{bmatrix} \begin{bmatrix} p(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t)$$

- First order ODE equation system:

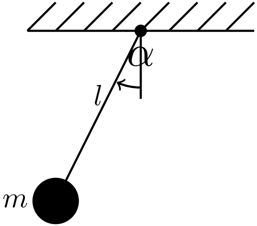
$$\dot{\boldsymbol{x}}(t) = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{\eta}{m} \end{bmatrix} \boldsymbol{x}(t) + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t)$$

- $\boldsymbol{x}(t) = \begin{bmatrix} p(t) & v(t) \end{bmatrix}^{\mathsf{T}}$ is the state of the system

Exercise: ODE to State-Space Representation

Tasks:

1. Find the (nonlinear) ODE for the pendulum below
2. Linearize the ODE using Taylor expansion around $\alpha = 0$
3. Write the linearized ODE as an ODE equation system



Hint: The moment of inertia for a pendulum is $I = ml^2$

Summary

- Separable models:

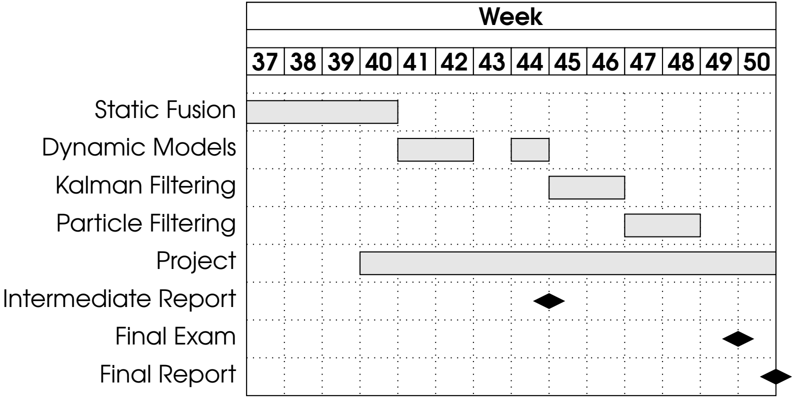
$$\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{\theta}_1)\boldsymbol{\theta}_2 + \boldsymbol{r}$$

- Cost function for the nonlinear parameters $\boldsymbol{\theta}_1$:

$$J(\boldsymbol{\theta}_1) = -\boldsymbol{y}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)(\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1))^{-1}\boldsymbol{G}(\boldsymbol{\theta}_1)^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{y}$$

- ODEs can be used to describe dynamic behavior
- Higher order ODEs can be reduced to first-order vector ODEs

Schedule


Week
37
38
39
40
41
42
43
44
45
46
47
48
49
50
Static Fusion
Dynamic Models
Kalman Filtering
Particle Filtering
Project
Intermediate Report
Final Exam
Final Report