CVXPY x NASA Course 2024

Philipp Schiele   Steven Diamond   Parth Nobel   Akshay Agrawal

July 1, 2024

Sensitivity Analysis and Robust Kalman Filters

Outline

Homework Review

Sensitivity

- Inequality constraints
- Equality constraints

Robust Kalman filtering

To the notebook

- `https://marimo.io/p/@cvxpy/optimal-trajectory-sol`
- The solution to 2(c) relies on background knowledge we did not cover

Challenge question

Background knowledge: At the minimum of a quadratic, its gradient is 0

- Differentiating objective with respect to t gives

$$\frac{\partial L}{\partial t}(c^\star, t^\star) = \sum_{i=1}^{m} 2\left(-\|u_i\|_2^2 + 2u_i^T c^\star + t^\star\right) = 0$$

Challenge question

Background knowledge: At the minimum of a quadratic, its gradient is 0

- Differentiating objective with respect to t gives

$$\frac{\partial L}{\partial t}(c^\star, t^\star) = \sum_{i=1}^{m} 2\left(-\|u_i\|_2^2 + 2u_i^T c^\star + t^\star\right) = 0$$

- Setting equal to zero and solving for $t^\star$

$$t^\star = \frac{1}{m}\sum_{i=1}^{m}\left(\|u_i\|_2^2 - 2u_i^T c^\star\right)$$

Challenge question

Background knowledge: At the minimum of a quadratic, its gradient is 0

- Differentiating objective with respect to t gives

$$\frac{\partial L}{\partial t}(c^\star, t^\star) = \sum_{i=1}^{m} 2\left(-\|u_i\|_2^2 + 2u_i^T c^\star + t^\star\right) = 0$$

- Setting equal to zero and solving for $t^\star$

$$t^\star = \frac{1}{m}\sum_{i=1}^{m}\left(\|u_i\|_2^2 - 2u_i^T c^\star\right)$$

- Adding $\|c^\star\|_2^2$ to both sides and recalling the expansion we had earlier,

$$t^\star + \|c^\star\|_2^2 = \frac{1}{m}\sum_{i=1}^{m}\|u_i - c^\star\|_2^2 \geq 0.$$

Outline

Questions to answer today

- Can we quantify how much a particular constraint matters to an optimization problem?
- When constraints have physical interpretation, can we learn more about the physical system?

Constraints

- Sometimes constraints in an optimization problem don't matter
-

$$\begin{array}{ll} \text{minimize} & |x-1| \\ \text{subject to} & x \geq -10 \end{array}$$

solution is $p^\star = 0, x^\star = 1$, regardless of the constraint

Constraints

- Sometimes constraints in an optimization problem don't matter
- $$\begin{array}{ll} \text{minimize} & |x-1| \\ \text{subject to} & x \geq -10 \end{array}$$

  solution is $p^\star = 0, x^\star = 1$, regardless of the constraint
- Often, contraints DO matter

  $$\begin{array}{ll} \text{minimize} & |x-1| \\ \text{subject to} & x \geq 10 \end{array}$$

  solution is $p^\star = 9, x^\star = 10$

Formal statement

- Let $p^\star(u)$ represent the optimal value of the following family of optimization problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq u \end{array}$$

- $p^\star : \mathbf{R} \to \mathbf{R}$ is a *function* of the perturbation and has units of the objective
- CVXPY computes

$$\frac{\mathrm{d}p^\star}{\mathrm{d}u}(0)$$

- These derivatives are meaningful

Example

- Words:

minimize average force on a 1 kg object as it travels in 1D

- Math (using trapezoidal integration for the ODE):

$$\text{minimize} \quad \frac{1}{n}\sum_{i=1}^{n} 1\,\text{kg} \times \frac{n}{3\,\text{s}}|v_i - v_{i-1}|$$

Example

- Words:

$$\begin{array}{ll} \text{minimize} & \text{average force on a } 1\,\text{kg object as it travels in 1D} \\ \text{subject to} & \text{from rest at position } 0\,\text{m} \end{array}$$

- Math (using trapezoidal integration for the ODE):

$$\begin{array}{ll} \text{minimize} & \frac{1}{n}\sum_{i=1}^{n} 1\,\text{kg} \times \frac{n}{3\,\text{s}} |v_i - v_{i-1}| \\ \text{subject to} & x_0 = 0\,\text{m},\, v_0 = 0\,\text{m}\,\text{s}^{-1} \end{array}$$

Example

- Words:

minimize	average force on a 1 kg object as it travels in 1D
subject to	from rest at position 0 m
	to rest at position 10 m

- Math (using trapezoidal integration for the ODE):

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{n}\sum_{i=1}^{n} 1\,\text{kg} \times \frac{n}{3\,\text{s}}|v_i - v_{i-1}| \\
\text{subject to} & x_0 = 0\,\text{m},\ v_0 = 0\,\text{m}\,\text{s}^{-1} \\
& x_n = 10\,\text{m},\ v_n = 0\,\text{m}\,\text{s}^{-1}
\end{array}
$$

Example

- Words:

$$\begin{array}{ll} \text{minimize} & \text{average force on a } 1\,\text{kg object as it travels in 1D} \\ \text{subject to} & \text{from rest at position } 0\,\text{m} \\ & \text{to rest at position } 10\,\text{m} \\ & \text{in } 3\,\text{s} \end{array}$$

- Math (using trapezoidal integration for the ODE):

$$\begin{array}{ll} \text{minimize} & \frac{1}{n}\sum_{i=1}^{n} 1\,\text{kg} \times \frac{n}{3\,\text{s}} |v_i - v_{i-1}| \\ \text{subject to} & x_0 = 0\,\text{m},\ v_0 = 0\,\text{m}\,\text{s}^{-1} \\ & x_n = 10\,\text{m},\ v_n = 0\,\text{m}\,\text{s}^{-1} \\ & x_i = x_{i-1} + \frac{3\,\text{s}}{2n}\left(v_i + v_{i-1}\right),\ i = 1, \dots, n \end{array}$$

Example

- Words:

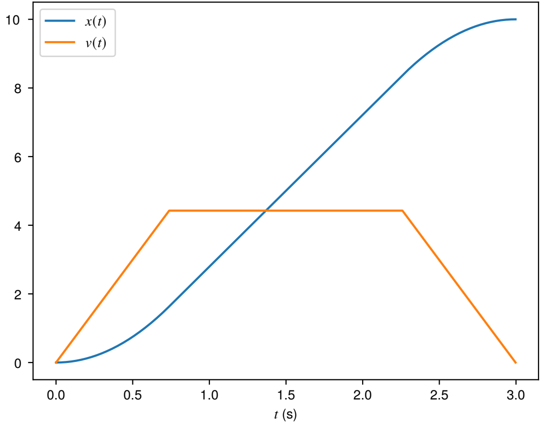
minimize	average force on a 1 kg object as it travels in 1D
subject to	from rest at position 0 m
	to rest at position 10 m
	in 3 s
	subject to an acceleration limit

- Math (using trapezoidal integration for the ODE):

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{n}\sum_{i=1}^{n} 1\,\text{kg} \times \frac{n}{3\,\text{s}}\left|v_i - v_{i-1}\right| \\
\text{subject to} & x_0 = 0\,\text{m},\, v_0 = 0\,\text{m}\,\text{s}^{-1} \\
& x_n = 10\,\text{m},\, v_n = 0\,\text{m}\,\text{s}^{-1} \\
& x_i = x_{i-1} + \frac{3\,\text{s}}{2n}\left(v_i + v_{i-1}\right),\, i = 1, \ldots, n \\
& \max_i \left(\frac{3\,\text{s}}{T}\left|v_i - v_{i-1}\right|\right) \leq 6\,\text{m}\,\text{s}^{-2} + u
\end{array}
$$

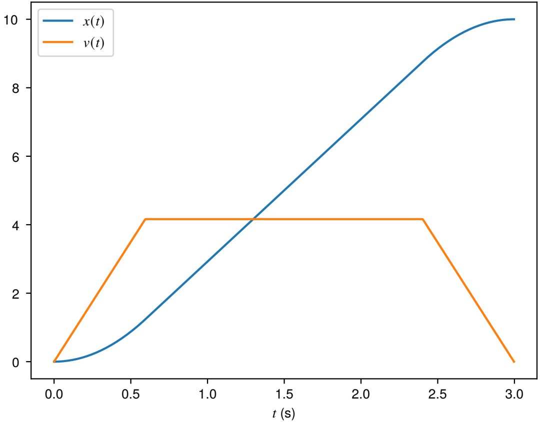
Example

Trajectory depends on u $u = 0\,\mathrm{m\,s^{-2}}$, $p^\star = 8.85\,\mathrm{N}$



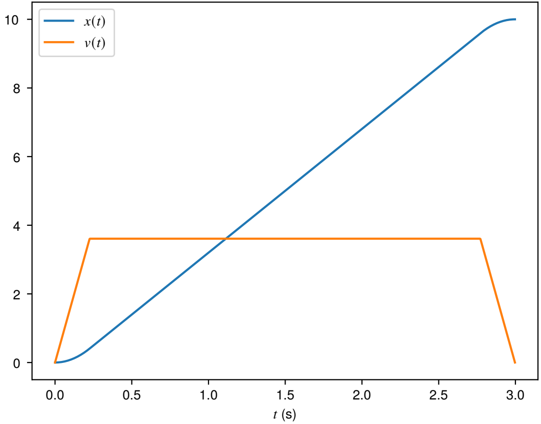
Example

Trajectory depends on u $\quad u = 1\,\mathrm{m\,s^{-2}},\ p^\star = 8.33\,\mathrm{N}$



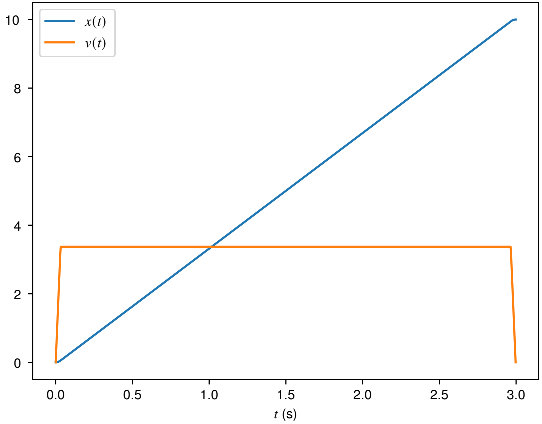
Example

Trajectory depends on u $\quad u = 10\,\mathrm{m\,s^{-2}}$, $p^\star = 7.22\,\mathrm{N}$



Example

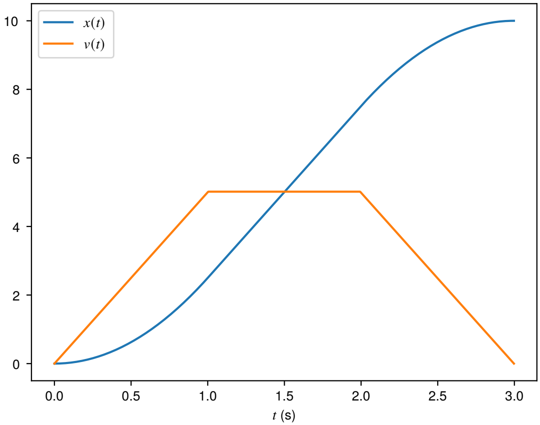
Trajectory depends on u $\quad u = 100\,\mathrm{m\,s^{-2}},\ p^\star = 6.75\,\mathrm{N}$



No impact on the trajectory

Example

Trajectory depends on u $\quad u = -1\,\mathrm{m\,s^{-2}}$, $p^\star = 10.03\,\mathrm{N}$



Dual variables

How does the average force change given small tweaks in u?

- We measure it with the derivative!
- $$\frac{\mathrm{d}p^\star}{\mathrm{d}u}(0)$$

  is called the **dual variable** of the acceleration constraint
- For inequality constraints, often denoted λ
- In our example, λ has units kg

Inequality constraints sometimes don't matter

- When an inequality constraint has no impact on an optimization problem, $\lambda = 0$
- Specifically, whenever an inequality constraint is not **tight**, $\lambda = 0$ (and other rare times too)
- **Complimentary slackness**

  At least one of λ or $f_1(x^\star)$ is equal to 0

- In our example,
  - $f_1(v) = \max_i \left(\frac{3\,\mathrm{s}}{T} |v_i - v_{i-1}| \right) - 6\,\mathrm{m\,s^{-2}}$
  - Complimentary slackness implies "if changing the acceleration limit doesn't change the cost of the optimal trajectory, then the mass hits the acceleration limit at at least one point in time"

Dual variables of equality constriants

- Nothing special like complimentary slackness
- They are just the derivative (when it exists) of the objective as you the equality constraint changes
- Interpreting the sign can be subtle

Outline

Homework Review

Sensitivity
- Inequality constraints
- Equality constraints

Robust Kalman filtering

The control/signal duality

- Imagine when we have a nice model for our system
- Controllers turn current state estimates into control signals
- Signal filters turn historical observations of a system and historical control inputs into state estimates

Problem setting

- Linear dynamical system

$$x_{t+1} = Ax_t + Bu_t, \qquad y_t = Cx_t$$

- Finite time horizon
- Controlled must be $u_t = F_t \hat{x}_t$
- Control problem is to pick u_t to control x_t given $\hat{x}_t$
- Signal problem is to find $\hat{x}_t$ to estimate x_t given historical y_t, u_t

Linear quadratic regulator (LQR) control

- Objective is to minimize

$$J\left(\{x_t\}_{t=0}^{T}, \{u_t\}_{t=0}^{T-1}\right) = x_T^T Q x_T + \sum_{t=0}^{T-1} (x_t^T Q x_t + u_t^T R u_t + 2 x_t^T N u_t)$$

  $Q, R \succeq 0$

- When u_t is unconstrained, the optimal solution is

$$F_t = (R + B^T P_{t+1} B)^{-1} (B^T P_{t+1} A + N^T)$$

- P_t given by the Riccati equations:

$$\begin{aligned} P_T &= Q \\ P_{t-1} &= A^T P_t A - (A^T P_t B + N)(R + B^T P_t B)^{-1}(B^T P_t A + N^T) + Q \end{aligned}$$

Kalman filtering

- we can estimate x_t for $t = 0, \ldots, N$ by solving the optimization problem

$$\begin{array}{ll} \text{minimize} & \sum_{t=0}^{N-1} \|w_t\|_2^2 + \tau \|v_t\|_2^2 \\ \text{subject to} & x_{t+1} = Ax_t + Bw_t \\ & y_t = Cx_t + v_t \end{array}$$

- τ is a regularization parameter
- this is a convex optimization problem

LQR + Kalman

- Optimality: The Kalman Filter + LQR is optimal for the model

$$
\begin{aligned}
x_{t+1} &= Ax_t + Bu_t + \epsilon_t \\
y_t &= Cx_t + \eta_t \\
u_t &= F_t \hat{x}_t \\
\epsilon_t &\sim \mathcal{N}(0, \Sigma_\epsilon) \\
\eta_t &\sim \mathcal{N}(0, \Sigma_\eta)
\end{aligned}
$$

with objective $\mathbb{E}[J]$, if $A, B, C, \Sigma_\epsilon, \Sigma_\eta$ are known

What about other models?

- We know it is optimal for:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + \epsilon_t \\ y_t &= Cx_t + \eta_t \\ u_t &= F_t \hat{x}_t \\ \epsilon_t &\sim \mathcal{N}(0, \Sigma_\epsilon) \\ \eta_t &\sim \mathcal{N}(0, \Sigma_\eta) \end{aligned}$$

  with objective $\mathbb{E}[J]$, if $A, B, C, \Sigma_\epsilon, \Sigma_\eta$ are known
- What about everything else?

Robust Kalman filtering

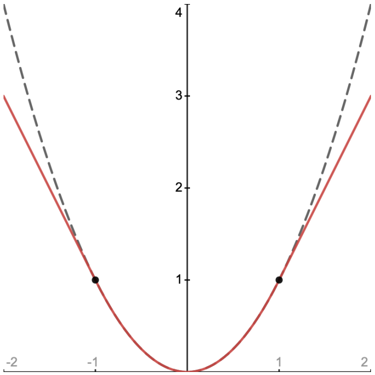
- the Kalman filter works well when v_t and w_t are Gaussian
- we can use a robust optimization approach to make the filter more robust to outliers
- using the Huber loss function, we can solve

$$\begin{array}{ll} \text{minimize} & \sum_{t=0}^{N-1} \phi_\epsilon(x_t) + \tau\phi_\eta(v_t) \\ \text{subject to} & x_{t+1} = Ax_t + Bw_t \\ & y_t = Cx_t + v_t, \end{array}$$

- where ϕ_ϵ encodes our understanding for the noise in the state transitions
- where ϕ_η encodes our understanding for the noise in the observations
- e.g. the rotationally-invariant Huber loss function

$$\phi^\rho(a) = \begin{cases} \|a\|_2^2 & \|a\|_2 \leq \rho \\ 2\rho\|a\|_2 - \rho^2 & \|a\|_2 > \rho \end{cases}$$

The Huber loss function



Penalty function approximation

$$\begin{array}{ll} \text{minimize} & \phi(r_1) + \cdots + \phi(r_m) \\ \text{subject to} & r = Ax - b \end{array}$$

($A \in \mathbf{R}^{m \times n}$, $\phi : \mathbf{R} \to \mathbf{R}$ is a convex penalty function)

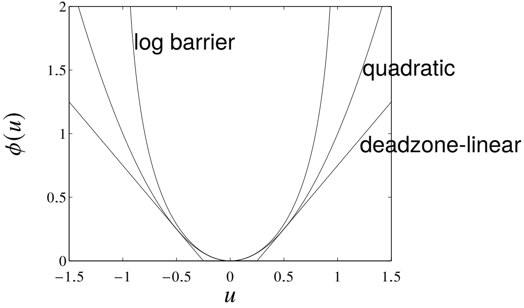
examples

- quadratic: $\phi(u) = u^2$
- deadzone-linear with width a:

$$\phi(u) = \max\{0, |u| - a\}$$

- log-barrier with limit a:

$$\phi(u) = \begin{cases} -a^2 \log(1 - (u/a)^2) & |u| < a \\ \infty & \text{otherwise} \end{cases}$$



Example: histograms of residuals

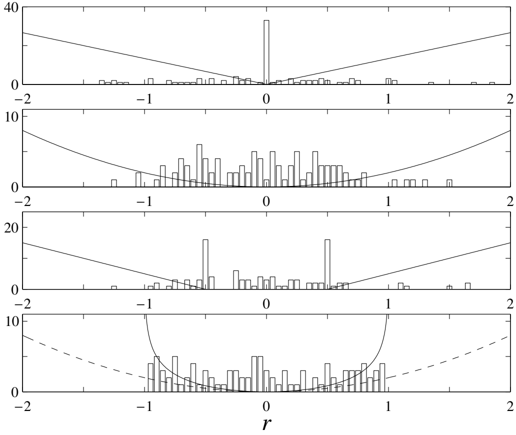
$A \in \mathbf{R}^{100 \times 30}$; shape of penalty function affects distribution of residuals

absolute value $\phi(u) = |u|$

square $\phi(u) = u^2$

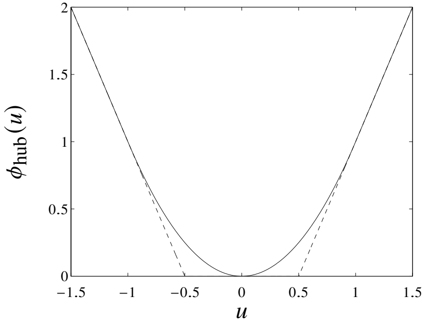
deadzone $\phi(u) = \max\{0, |u| - 0.5\}$

log-barrier $\phi(u) = -\log(1 - u^2)$



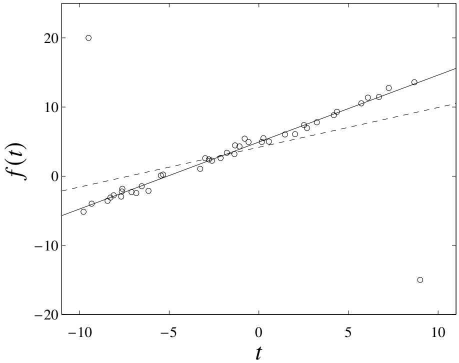
Huber penalty function

$$\phi_{\text{hub}}(u) = \begin{cases} u^2 & |u| \leq M \\ M(2|u| - M) & |u| > M \end{cases}$$



- linear growth for large u makes approximation less sensitive to outliers
- called a **robust penalty**

Example



- 42 points (circles) t_i, y_i, with two outliers
- affine function $f(t) = \alpha + \beta t$ fit using quadratic (dashed) and Huber (solid) penalty

Weird noise model? Can you still regularize for it?

- Many assumptions about how the data should look can be imposed
- For example, what if we know ϵ is so strongly autocorrelated that it should be piecewise constant
- We can impose $\|Dv\|_1$ regularization

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

  imposing first-order sparsity
- Expecting noise is being introduced in a low-dimensional linear space? Use the nuclear norm