14. Constrained optimization

- equality constrained optimization
- penalty method
- augmented Lagrangian method
- constrained nonlinear least squares
- nonlinear control example

Equality constrained optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

- $f : \mathbb{R}^n \to \mathbb{R}$; $g_i : \mathbb{R}^n \to \mathbb{R}$
- we let $g(x) = (g_1(x), \dots, g_p(x))$
- a point x satisfying $g(x) = 0$ is called a *feasible point*
- $\hat{x}$ is a solution if it is feasible and $f(\hat{x}) \leq f(x)$ for all feasible x

Regular point: a feasible point x is a *regular point* if the vectors

$$\nabla g_1(x),\ \nabla g_2(x),\ \dots,\ \nabla g_p(x)$$

are linearly independent

Lagrangian function

the *Lagrangian function* is defined as

$$L(x, z) = f(x) + \sum_{i=1}^{p} z_i g_i(x)$$

- $z = (z_1, \ldots, z_p)$ is a p-vector
- the entries of z_i are called the *Lagrange multipliers*
- the *gradient of Lagrangian* is

  $$\nabla L(x, z) = \begin{bmatrix} \nabla_x L(x, z) \\ \nabla_z L(x, z) \end{bmatrix}$$

  where

  $$\begin{aligned} \nabla_x L(x, z) &= \nabla f(x) + \sum_{i=1}^{p} z_i \nabla g_i(x) \\ \nabla_z L(x, z) &= g(x) \end{aligned}$$

Method of Lagrange multipliers

if $\hat{x}$ is a regular point and a local minimizer, then there exists a vector $\hat{z}$ such that

$$\nabla_x L(\hat{x}, \hat{z}) = \nabla f(\hat{x}) + \sum_{i=1}^{p} \hat{z}_i \nabla g_i(\hat{x}) = 0$$
$$\nabla_z L(\hat{x}, \hat{z}) = g(\hat{x}) = 0$$

- $\hat{z}$ is called an *optimal Lagrange multiplier*
- these are necessary conditions but not sufficient
- there can be points (x, z) satisfying the above but x is not a local minimizer
- the above method is known as the *method of Lagrange multipliers*
- called *KKT conditions* or *Lagrange conditions*

Example

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1^2 + 2x_2^2 = 1 \end{array}$$

- the Lagrangian is

$$L(x, z) = x_1^2 + x_2^2 + z(x_1^2 + 2x_2^2 - 1)$$

- the necessary optimality conditions are

$$\nabla_x L(x, z) = \begin{bmatrix} 2x_1 + 2x_1 z \\ 2x_2 + 4x_2 z \end{bmatrix} = 0$$

$$\nabla_z L(x, z) = x_1^2 + 2x_2^2 - 1 = 0$$

- solving, we get the stationary points

$$x = (0, \pm\tfrac{1}{\sqrt{2}}), \quad z = -1/2$$

and

$$x = (\pm 1, 0), \quad z = -1$$

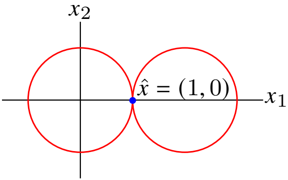
- all feasible points are regular since $\nabla g(x) = (2x_1, 4x_2)$ is linearly independent
- thus, any minimizer to the above problem must satisfy the optimality conditions
- checking the value of the objective, we see that it is smallest at

$$x^{(1)} = (0, \tfrac{1}{\sqrt{2}}) \quad \text{and} \quad x^{(2)} = (0, -\tfrac{1}{\sqrt{2}})$$

- therefore, the points $x^{(1)}$ and $x^{(2)}$ are candidate minimizers

Example

$$
\begin{array}{ll}
\text{minimize} & x_2 \\
\text{subject to} & x_1^2 + x_2^2 = 1 \\
 & (x_1 - 2)^2 + x_2^2 = 1
\end{array}
$$



one feasible point $\hat{x} = (1, 0)$, thus optimal

- $\hat{x}$ is not regular as $\nabla g_1(\hat{x}) = (2, 0)$, $\nabla g_2(\hat{x}) = (-2, 0)$ are dependent
- the Lagrangian is

$$L(x, z) = x_2 + z_1(x_1^2 + x_2^2 - 1) + z_2\left((x_1 - 2)^2 + x_2^2 - 1\right)$$

- the necessary condition

$$\nabla_x L(x, z) = \begin{bmatrix} 2x_1 z_1 + 2(x_1 - 2)z_2 \\ 1 + 2x_2(z_1 + z_2) \end{bmatrix} = 0$$

cannot be satisfied at $\hat{x} = (1, 0)$

Outline

- equality constrained optimization
- **penalty method**
- augmented Lagrangian method
- constrained nonlinear least squares
- nonlinear control example

Quadratic penalty formulation

$$\text{minimize} \quad f(x) + \mu \|g(x)\|^2$$

- $g(x) = (g_1(x), \ldots, g_p(x))$
- $\mu \in \mathbb{R}$ is the *penalty parameter*
- $\mu \|g(x)\|^2 = \mu \sum_{i=1}^{p} (g_i(x))^2$ penalize constraints violation
- a solution of the above problem might not feasible
- for large μ we expect to have small values $\big(g_i(x)\big)^2$
- minimizing the above for an increasing sequence μ is called the *penalty method*

Penalty method

given a starting point $x^{(1)}$, $\mu^{(1)}$, and a solution tolerance $\epsilon > 0$

repeat for $k = 1, 2, \ldots$

1. set $x^{(k+1)}$ to be the (approximate) solution to

$$\text{minimize} \quad f(x) + \mu^{(k)} \|g(x)\|^2$$

   using an unconstrained optimization method with initial point $x^{(k)}$

2. update $\mu^{(k+1)} = 2\mu^{(k)}$

- penalty method is terminated when $\|g(x^{(k)})\|$ becomes sufficiently small
- simple and easy to implement
- feasibility $g(x^{(k)}) = 0$ is only satisfied approximately for $\mu^{(k-1)}$ large enough
- $\mu^{(k)}$ increases rapidly and must become large to drive $g(x)$ to (near) zero
- for large $\mu^{(k)}$, step 1 can take a large number of iterations, or fail

Connection to optimality condition

recall optimality condition

$$\nabla f(\hat{x}) + Dg(\hat{x})^T \hat{z} = 0, \quad g(\hat{x}) = 0$$

- $x^{(k+1)}$ satisfies optimality condition for unconstrained problem:

$$\nabla f(x^{(k+1)}) + 2\mu^{(k)} Dg(x^{(k+1)})^T g(x^{(k+1)}) = 0$$

- if we define $z^{(k+1)} = 2\mu^{(k)} g(x^{(k+1)})$, this can be written as

$$\nabla f(x^{(k+1)}) + Dg(x^{(k+1)})^T z^{(k+1)} = 0$$

- so $x^{(k+1)}$ and $z^{(k+1)}$ satisfy first equation in KKT optimality condition

- feasibility $g(x^{(k+1)}) = 0$ is only satisfied approximately for $\mu^{(k)}$ large enough

- feasibility $g(x^{(k+1)}) = 0$ holds in the limit

Outline

- equality constrained optimization
- penalty method
- **augmented Lagrangian method**
- constrained nonlinear least squares
- nonlinear control example

Minimizing the Lagrangian

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & g(x) = 0
\end{array}
$$

- $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^p \to \mathbb{R}$
- Lagrangian: $L(x, z) = f(x) + z^T g(x)$ where $z \in \mathbb{R}^p$
- problem is equivalent to (for any z)

$$
\begin{array}{ll}
\text{minimize} & L(x, z) = f(x) + z^T g(x) \\
\text{subject to} & g(x) = 0
\end{array}
$$

- if $\hat{x}$ is a solution and a regular point, then

$$
\nabla_x L(\hat{x}, \hat{z}) = 0 \quad \text{for some } \hat{z}
$$

Augmented Lagrangian

the **augmented Lagrangian** (AL) is

$$\begin{aligned} L_\mu(x,z) &= L(x,z) + \mu\|g(x)\|^2 \\ &= f(x) + z^T g(x) + \mu\|g(x)\|^2 \end{aligned}$$

- this is the Lagrangian $L(x,z)$ augmented with a quadratic penalty
- μ is a positive penalty parameter
- augmented Lagrangian is the Lagrangian of the equivalent problem

$$\begin{array}{ll} \text{minimize} & f(x) + \mu\|g(x)\|^2 \\ \text{subject to} & g(x) = 0 \end{array}$$

- solution of the original problem is also a solution of the AL formulation
- AL method minimizes $L_\mu(x,z)$ for a sequence of values of z and μ

Lagrange multiplier update

- minimizer $\tilde{x}$ of augmented Lagrangian $L_\mu(x, z)$ satisfies

$$\nabla f(\tilde{x}) + Dg(\tilde{x})^T(2\mu g(\tilde{x}) + z) = 0$$

- if we define $\tilde{z} = z + 2\mu g(\tilde{x})$ this can be written as

$$\nabla f(\tilde{x}) + Dg(\tilde{x})^T \tilde{z} = 0$$

- this is the first equation in the optimality conditions

$$\nabla f(\hat{x}) + Dg(\hat{x})^T \hat{z} = 0, \quad g(\hat{x}) = 0$$

- shows that if $g(\tilde{x}) = 0$, then $\tilde{x}$ is optimal
- if $g(\tilde{x})$ is not small, suggests $\tilde{z}$ is a good update for z

Augmented Lagrangian algorithm

given $x^{(1)}$, $z^{(1)}$, $\mu^{(1)}$, and a solution tolerance $\epsilon > 0$

repeat for $k = 1, 2, \ldots$

1. set $x^{(k+1)}$ to be an (approximate) solution to

$$\text{minimize} \quad f(x) + (z^{(k)})^T g(x) + \mu^{(k)} \|g(x)\|^2$$

   using any unconstrained optimization method with initial point $x^{(k)}$

2. update $z^{(k)}$:

$$z^{(k+1)} = z^{(k)} + 2\mu^{(k)} g(x^{(k+1)})$$

3. set $\mu^{(k)}$ as constant or

$$\begin{cases} \mu^{(k)} & \text{if} \quad \|g(x^{(k+1)})\| < 0.25 \|g(x^{(k)})\| \\ 2\mu^{(k)} & \text{otherwise} \end{cases}$$

- μ is increased only when needed, more slowly than in penalty method
- continues until $g(x^{(k)})$ and/or $\nabla L(x^{(k)}, z^{(k)})$ are sufficiently small

Example

$$\begin{array}{ll} \text{minimize} & e^{3x_1} + e^{-4x_2} \\ \text{subject to} & x_1^2 + x_2^2 = 1 \end{array}$$

the augmented Lagrangian function is:

$$L_\mu(x, z) = e^{3x_1} + e^{-4x_2} + z\left(x_1^2 + x_2^2 - 1\right) + \mu(x_1^2 + x_2^2 - 1)^2$$

- initial points $x^{(1)} = (0, 0)$ and $z^{(1)} = -1$, and $\mu^{(k)} = 10$
- for the inner minimization problems we use Newton's method with stepsize $t = 1$:

$$\hat{x} \leftarrow \hat{x} - t\nabla^2 L_\mu(\hat{x}, z^{(k)})^{-1} \nabla L_\mu(\hat{x}, z^{(k)})$$

the gradient and Hessian are:

$$\nabla_x L_\mu(x,z) = \begin{bmatrix} 3e^{3x_1} + 2zx_1 + 4\mu x_1(x_1^2 + x_2^2 - 1) \\ -4e^{-4x_2} + 2zx_2 + 4\mu x_2(x_1^2 + x_2^2 - 1) \end{bmatrix}$$

and

$$\nabla_x^2 L_\mu(x,z) = \begin{bmatrix} 9e^{3x_1} + 2z + 4\mu(x_1^2 + x_2^2 - 1) + 8\mu x_1^2 & 8\mu x_1 x_2 \\ 8\mu x_1 x_2 & 16e^{-4x_2} + 2z + 4\mu(x_1^2 + x_2^2 - 1) + 8\mu x_2^2 \end{bmatrix}$$

Newton method starts from $\hat{x} = x^{(k)}$ and and stops if $\|\nabla L_\mu(\hat{x}, z^{(k)})\| < 10^{-5}$

the value $x^{(k+1)}$ is then set to $\hat{x}$ and the Lagrange multiplier is subsequently updated:

$$z^{(k+1)} = z^{(k)} + 2\mu\big((x_1^{(k+1)})^2 + (x_2^{(k+1)})^2 - 1\big)$$

after executing the augmented Lagrangian method until $\|g(x^{(k)})\| < 10^{-6}$ or 50 iterations, the results are approximately $\hat{x} = (-0.7483, 0.6633)$ and $z^\star = 0.2123$

MATLAB code implementation

```
mu=10;
%% AL gradient and Hessian
g=@(x,z)[3*exp(3*x(1))+2*z*x(1)+4*mu*x(1)*(x(1)^2+x(2)^2-1);
-4*exp(-4*x(2))+2*z*x(2)+4*mu*x(2)*(x(1)^2+x(2)^2-1)];
hess=@(x,z)[9*exp(3*x(1))+2*z+4*mu*(x(1)^2+x(2)^2-1)+...
8*mu*x(1)^2 8*mu*x(1)*x(2);
8*mu*x(1)*x(2)  16*exp(-4*x(2))+2*z+4*mu*(x(1)^2+x(2)^2-1)+8*mu*x(2)^2];
%% AL method
x=[0;0];
z=-1;
for i=1:50
% Newton inner minimization
while (norm(g(x,z)) >= 1e-5)
d = -hess(x,z)\g(x,z);
x = x+d;
end
% Lagrange update
z=z+mu*(x(1)^2+x(2)^2-1);
if norm((x(1)^2+x(2)^2-1))<1e-6
return
end
end
```

Outline

- equality constrained optimization
- penalty method
- augmented Lagrangian method
- **constrained nonlinear least squares**
- nonlinear control example

Constrained nonlinear least squares

$$\begin{array}{ll}\text{minimize} & f_1(x)^2 + \cdots + f_p(x)^2 \\ \text{subject to} & g_1(x) = 0, \ldots, g_p(x) = 0\end{array}$$

in vector notation:

$$\begin{array}{ll}\text{minimize} & \|f(x)\|^2 \\ \text{subject to} & g(x) = 0\end{array}$$

with $f(x) = (f_1(x), \ldots, f_m(x))$, $g(x) = (g_1(x), \ldots, g_p(x))$

- $f_i(x)$ is ith (scalar) *residual*; $g_i(x) = 0$ is ith (scalar) *equality constraint*
- x is feasible if it satisfies the constraints $g(x) = 0$
- $\hat{x}$ is a solution if it is feasible and $\|f(x)\|^2 \geq \|f(\hat{x})\|^2$ for all feasible x

Lagrange multipliers

the **Lagrangian** of the problem is the function

$$\begin{aligned} L(x, z) &= \|f(x)\|^2 + z_1 g_1(x) + \cdots + z_m g_m(x) \\ &= \|f(x)\|^2 + g(x)^T z \end{aligned}$$

- p-vector $z = (z_1, \ldots, z_p)$ is vector of *Lagrange multipliers*
- gradient of Lagrangian with respect to x is

$$\nabla_x L(\hat{x}, \hat{z}) = 2Df(\hat{x})^T f(\hat{x}) + Dg(\hat{x})^T \hat{z}$$

- gradient with respect to z is

$$\nabla_z L(\hat{x}, \hat{z}) = g(\hat{x})$$

Optimality condition: if $\hat{x}$ is optimal, then there exists $\hat{z}$ such that

$$2Df(\hat{x})^T f(\hat{x}) + Dg(\hat{x})^T \hat{z} = 0, \quad g(\hat{x}) = 0$$

provided the rows of $Dg(\hat{x})$ are linearly independent

Constrained (linear) least squares

$$\begin{array}{ll} \text{minimize} & (1/2)\|Ax - b\|^2 \\ \text{subject to} & Cx = d \end{array}$$

- a special case of the nonlinear problem with $f(x) = Ax - b, g(x) = Cx - d$
- apply general optimality condition:

$$\begin{aligned} Df(\hat{x})^T f(\hat{x}) + Dg(\hat{x})^T \hat{z} = A^T(A\hat{x} - b) + C^T \hat{z} &= 0 \\ g(\hat{x}) = C\hat{x} - d &= 0 \end{aligned}$$

- these are the KKT equations

$$\left[\begin{array}{cc} A^TA & C^T \\ C & 0 \end{array}\right] \left[\begin{array}{c} \hat{x} \\ \hat{z} \end{array}\right] = \left[\begin{array}{c} A^Tb \\ d \end{array}\right]$$

Penalty algorithm

given a starting point $x^{(1)}$, $\mu^{(1)}$, and solution tolerance ϵ

repeat for $k = 1, 2, \ldots$

1. set $x^{(k+1)}$ to be an (approximate) solution to

$$\text{minimize} \quad \left\| \begin{bmatrix} f(x) \\ \sqrt{\mu} g(x) \end{bmatrix} \right\|^2$$

   using the Levenberg-Marquardt algorithm, starting from initial point $x^{(k)}$

2. update $\mu^{(k)} = 2\mu^{(k)}$

   if stopping criteria holds, stop and output $x^{(k+1)}$

Augmented Lagrangian

the **augmented Lagrangian** for the constrained NLLS problem is

$$
\begin{aligned}
L_\mu(x,z) &= L(x,z) + \mu\|g(x)\|^2 \\
&= \|f(x)\|^2 + g(x)^T z + \mu\|g(x)\|^2
\end{aligned}
$$

- equivalent expressions for augmented Lagrangian

$$
\begin{aligned}
L_\mu(x,z) &= \|f(x)\|^2 + g(x)^T z + \mu\|g(x)\|^2 \\
&= \|f(x)\|^2 + \mu\left\|g(x) + \tfrac{1}{2\mu}z\right\|^2 - \frac{1}{2\mu}\|z\|^2 \\
&= \left\|\begin{bmatrix} f(x) \\ \sqrt{\mu}g(x) + z/(2\sqrt{\mu}) \end{bmatrix}\right\|^2 - \frac{1}{2\mu}\|z\|^2
\end{aligned}
$$

- can be minimized over x (for fixed μ, z) by Levenberg-Marquardt method:

$$
\text{minimize} \quad \left\|\begin{bmatrix} f(x) \\ \sqrt{\mu}g(x) + z/(2\sqrt{\mu}) \end{bmatrix}\right\|^2
$$

Augmented Lagrangian algorithm

given: $z^{(1)} = 0$, $\mu^{(1)} = 1$, and $x^{(1)}$

repeat for $k = 1, 2 \ldots$

1. set $x^{(k+1)}$ to be the (approximate) solution of

$$\text{minimize} \quad \left\| \begin{bmatrix} f(x) \\ \sqrt{\mu^{(k)}} g(x) + z^{(k)} / (2\sqrt{\mu^{(k)}}) \end{bmatrix} \right\|^2$$

   using Levenberg-Marquardt algorithm, starting from initial point $x^{(k)}$

2. multiplier update:

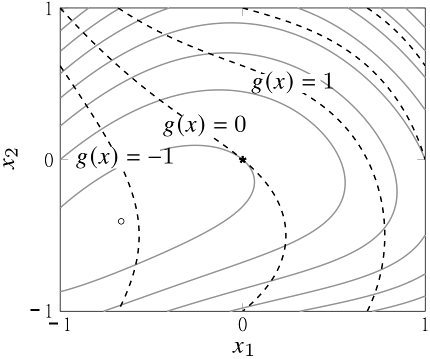
$$z^{(k+1)} = z^{(k)} + 2\mu^{(k)} g(x^{(k+1)})$$

3. penalty parameter update:

$$\mu^{(k+1)} = \begin{cases} \mu^{(k)} & \text{if } \|g(x^{(k+1)})\| < 0.25 \|g(x^{(k)})\| \\ \mu^{(k+1)} = 2\mu^{(k)} & \text{otherwise} \end{cases}$$

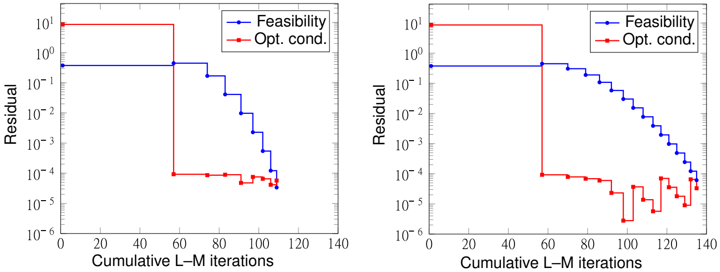
Example

$$f(x_1, x_2) = \left[\begin{array}{c} x_1 + \exp(-x_2) \\ x_1^2 + 2x_2 + 1 \end{array} \right], \quad g(x_1, x_2) = x_1 + x_1^3 + x_2 + x_2^2$$

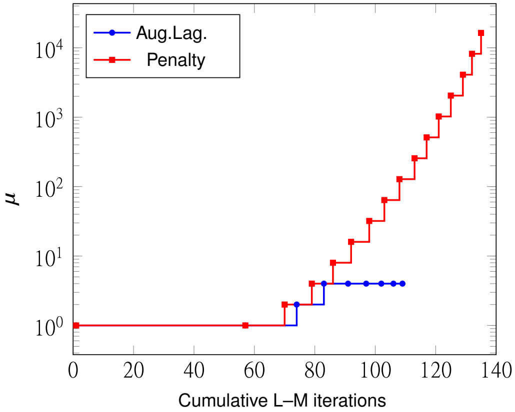


- solid: contour lines of $\|f(x)\|^2$
- dashed: contour lines of $g(x)$
- $*$: solution $\hat{x} = (0, 0)$

Convergence



- left: augmented Lagrangian, right: penalty method
- blue curve is norm $\|g(x^{(k)})\|$
- red curve is norm of $\|2Df(x^{(k)})^T f(x^{(k)}) + Dg(x^{(k)})^T z^{(k)}\|$



Aug.Lag.
Penalty
10^4
10^3
10^2
10^1
10^0
μ
0
20
40
60
80
100
120
140
Cumulative L–M iterations

Outline

- equality constrained optimization
- penalty method
- augmented Lagrangian method
- constrained nonlinear least squares
- **nonlinear control example**

Nonlinear dynamical system

a *nonlinear dynamical system* has the form

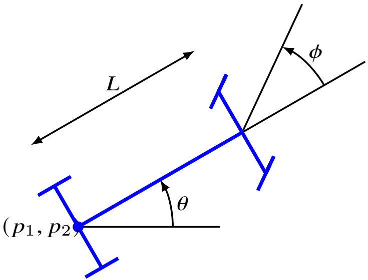
$$x_{k+1} = f(x_k, u_k), \quad k = 1, \ldots, K$$

- $x_k \in \mathbb{R}^n$ is the *state vector* at instant k
- $u_k \in \mathbb{R}^m$ is the *input* or *control* at instant k
- $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$ describes evolution of the system (system dynamics)
- examples: vehicle dynamics, robots, chemical plants evolution...

Optimal control

- initial state $x_1 = x_{\text{initial}}$ is known
- choose the inputs $u_1, \ldots, u_K$ to achieve some goal for the states/inputs

Simple model of a car



$$\frac{dp_1}{dt} = s(t)\cos\theta(t)$$
$$\frac{dp_2}{dt} = s(t)\sin\theta(t)$$
$$\frac{d\theta}{dt} = \frac{s(t)}{L}\tan\phi(t)$$

- L wheelbase (length)
- $p(t)$ position
- $\theta(t)$ orientation (angle)
- $\phi(t)$ steering angle
- $s(t)$ speed
- we control speed s and steering angle ϕ

Discretized car dynamics

$$
\begin{aligned}
p_1(t+h) &\approx p_1(t) + hs(t)\cos\theta(t) \\
p_2(t+h) &\approx p_2(t) + hs(t)\sin\theta(t \\
\theta(t+h) &\approx \theta(t) + h(s(t)/L)\tan\phi(t)
\end{aligned}
$$

- h is a small time interval
- let state vector $x_k = (p_1(kh), p_2(kh), \theta(kh))$
- let input vector $u_k = (s(kh), \phi(kh))$
- discretized model $x_{k+1} = f\,(x_k, u_k)$ with

$$
h\,(x_k, u_k) = x_k + h\,(u_k)_1 \begin{bmatrix} \cos\,(x_k)_3 \\ \sin\,(x_k)_3 \\ (\tan\,(u_k)_2)\,/L \end{bmatrix}
$$

Car control problem

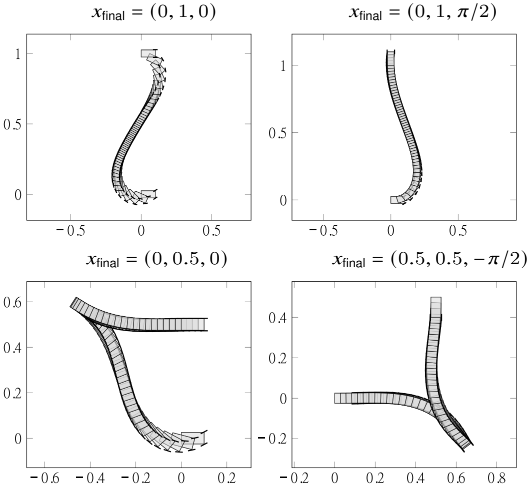
- move car from given initial to desired final position and orientation
- using a small and slowly varying input sequence

Problem formulation

$$\begin{array}{ll} \text{minimize} & \sum_{k=0}^{K} \|u_k\|^2 + \gamma \sum_{k=0}^{K-1} \|u_{k+1} - u_k\|^2 \\ \text{subject to} & x_2 = f(0, u_1) \\ & x_{k+1} = f(x_k, u_k), \quad k = 2, \dots, K-1 \\ & x_{\text{final}} = f(x_K, u_K) \end{array}$$

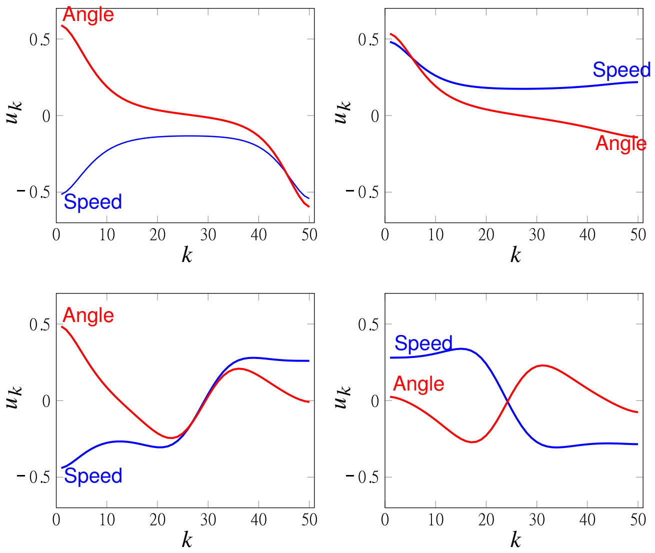
- variables $u_1, \dots, u_K$, and $x_2, \dots, x_K$
- the initial state is assumed to be zero
- the objective ensures the input is small with little variation
- $\gamma > 0$ is an input variation trade-off parameter

Four solution trajectories



solution trajectories with different final states; the outline of the car shows the position $(p_1(kh); p_2(kh))$, orientation $\theta(kh)$, and the steering angle $\phi(kh)$ at time kh

Inputs for four trajectories



References and further readings

- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares,* Cambridge University Press, 2018.
- L. Vandenberghe. *EE133A lecture notes,* Univ. of California, Los Angeles. (`http://www.seas.ucla.edu/~vandenbe/ee133a.html`)