19. Constrained nonlinear least squares

Outline

Constrained nonlinear least squares

Penalty method

Augmented Lagrangian method

Nonlinear control example

Constrained nonlinear least squares

- add equality constraints to nonlinear least squares problem:

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

- $f_i(x)$ is ith (scalar) *residual*; $g_i(x) = 0$ is ith (scalar) equality constraint
- with vector notation $f(x) = (f_1(x), \ldots, f_m(x))$, $g(x) = (g_1(x), \ldots, g_p(x))$

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

- 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
- problem is difficult to solve in general, but useful heuristics exist

Lagrange multipliers

- the *Lagrangian* of the problem is the function

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

- p-vector $z = (z_1, \ldots, z_p)$ is vector of *Lagrange multipliers*
- method of Lagrange multipliers: if $\hat{x}$ is a solution, then there exists $\hat{z}$ with

$$
\frac{\partial L}{\partial x_i}(\hat{x},\hat{z}) = 0, \quad i = 1,\ldots,n. \qquad \frac{\partial L}{\partial z_i}(\hat{x},\hat{z}) = 0, \quad i = 1,\ldots,p
$$

  (provided the gradients $\nabla g_1(\hat{x}), \ldots, \nabla g_p(\hat{x})$ are linearly independent)
- $\hat{z}$ is called an *optimal Lagrange multiplier*

Optimality condition

- 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

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

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

- this condition is necessary for optimality but not sufficient

Constrained (linear) least squares

- recall constrained least squares problem

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

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

$$2Df(\hat{x})^T f(\hat{x}) + Dg(\hat{x})^T \hat{z} = 2A^T(A\hat{x} - b) + C^T\hat{z} = 0, \qquad g(\hat{x}) = C\hat{x} - d = 0$$

- these are the KKT equations

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

Outline

Constrained nonlinear least squares

Penalty method

Augmented Lagrangian method

Nonlinear control example

Penalty method

- solve sequence of (unconstrained) nonlinear least squares problems

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

- μ is a positive *penalty parameter*
- instead of insisting on $g(x) = 0$ we assign a penalty to deviations from zero
- for increasing sequence $\mu^{(1)}, \mu^{(2)}, \ldots$, compute $x^{(k+1)}$ by minimizing

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

- $x^{(k+1)}$ is computed by Levenberg–Marquardt algorithm started at $x^{(k)}$

Termination

- recall optimality condition

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

- $x^{(k)}$ satisfies normal equations for linear least squares problem:

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

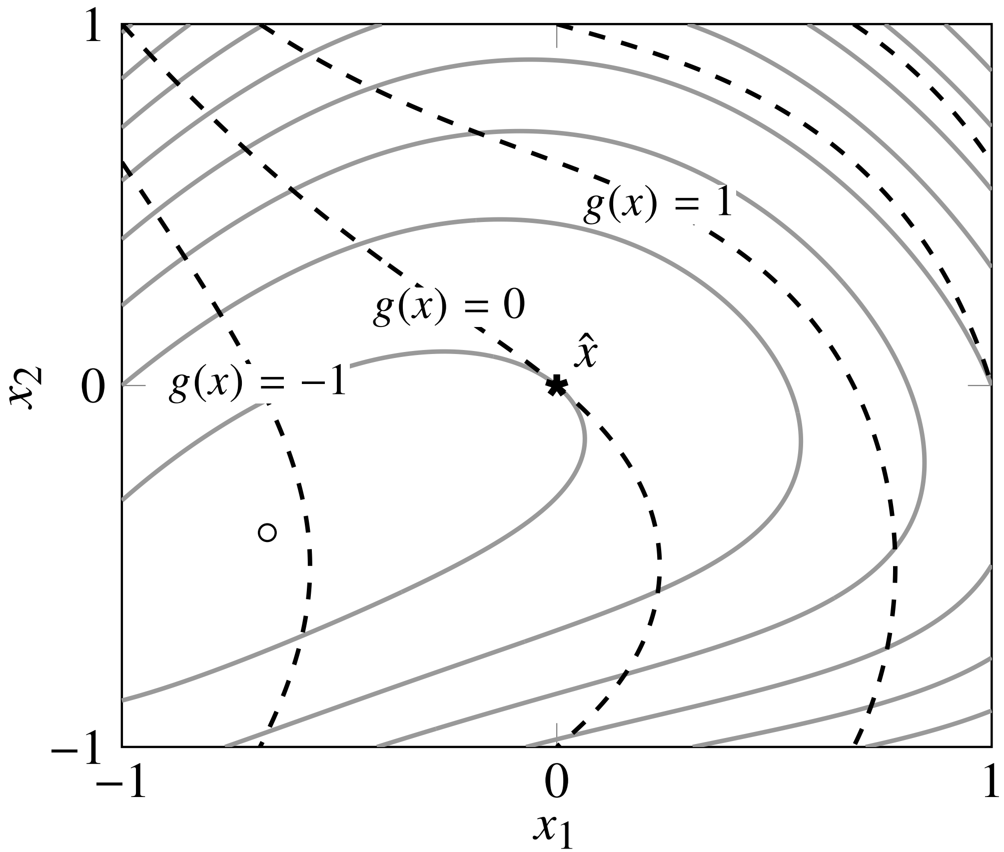
- if we define $z^{(k)} = 2\mu^{(k-1)} g(x^{(k)})$, this can be written as

$$2Df(x^{(k)})^T f(x^{(k)}) + Dg(x^{(k)})^T z^{(k)} = 0$$

- we see that $x^{(k)}$, $z^{(k)}$ satisfy first equation in optimality condition
- feasibility $g(x^{(k)}) = 0$ is only satisfied approximately for $\mu^{(k-1)}$ large enough
- penalty method is terminated when $\|g(x^{(k)})\|$ becomes sufficiently small

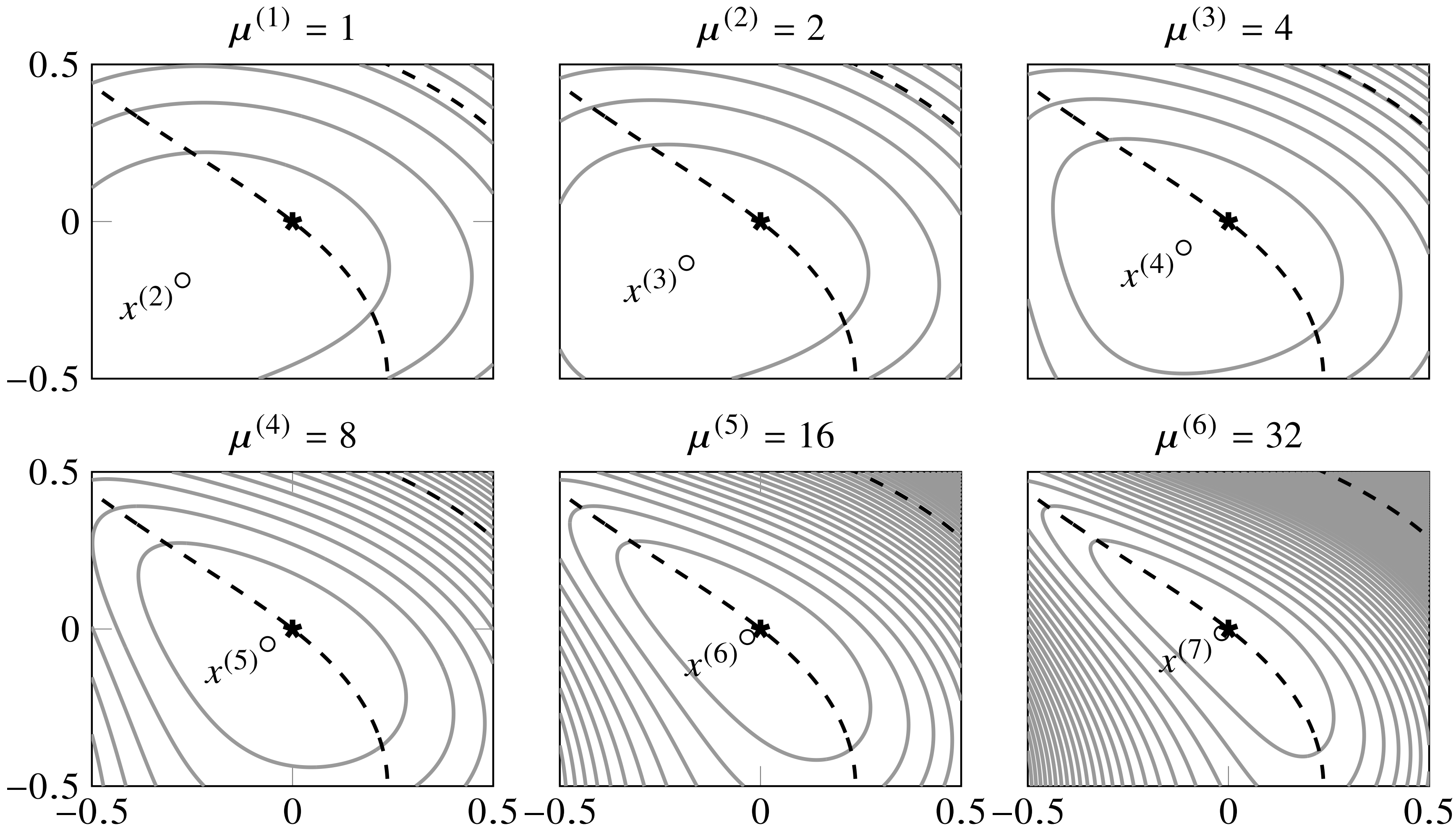
Example

$$f(x_1, x_2) = \begin{bmatrix} x_1 + \exp(-x_2) \\ x_1^2 + 2x_2 + 1 \end{bmatrix}, \qquad g(x_1, x_2) = x_1 + x_1^3 + x_2 + x_2^2$$



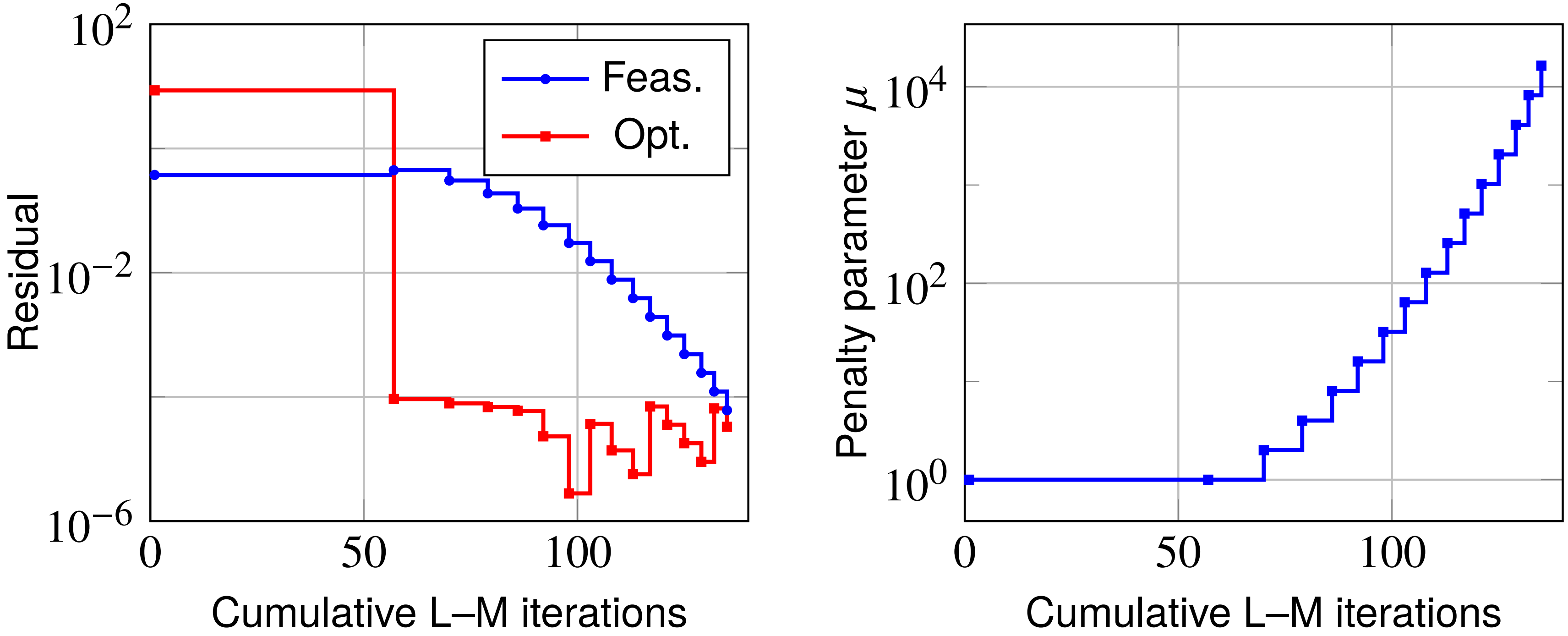
- solid: contour lines of $\|f(x)\|^2$
- dashed: contour lines of $g(x)$
- $\hat{x}$ is solution

First six iterations



solid lines are contour lines of $\|f(x)\|^2 + \mu^{(k)}\|g(x)\|^2$

Convergence



- figure on the left shows residuals in optimality condition
- blue curve is norm of $g(x^{(k)})$
- red curve is norm of $2Df(x^{(k)})^T f(x^{(k)}) + Dg(x^{(k)})^T z^{(k)}$

Outline

Constrained nonlinear least squares

Penalty method

Augmented Lagrangian method

Nonlinear control example

Drawback of penalty method

- $\mu^{(k)}$ increases rapidly and must become large to drive $g(x)$ to (near) zero
- for large $\mu^{(k)}$, nonlinear least squares subproblem becomes harder
- for large $\mu^{(k)}$, Levenberg–Marquardt method can take a large number of iterations, or fail

Augmented Lagrangian

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

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

- 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

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

Minimizing augmented Lagrangian

- equivalent expressions for augmented Lagrangian

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

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

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

Lagrange multiplier update

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

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

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

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

- this is the first equation in the optimality conditions

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

- 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

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

$$\|f(x)\|^2 + \mu^{(k)}\|g(x) + z^{(k)}/(2\mu^{(k)})\|^2$$

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

2. *multiplier update:*

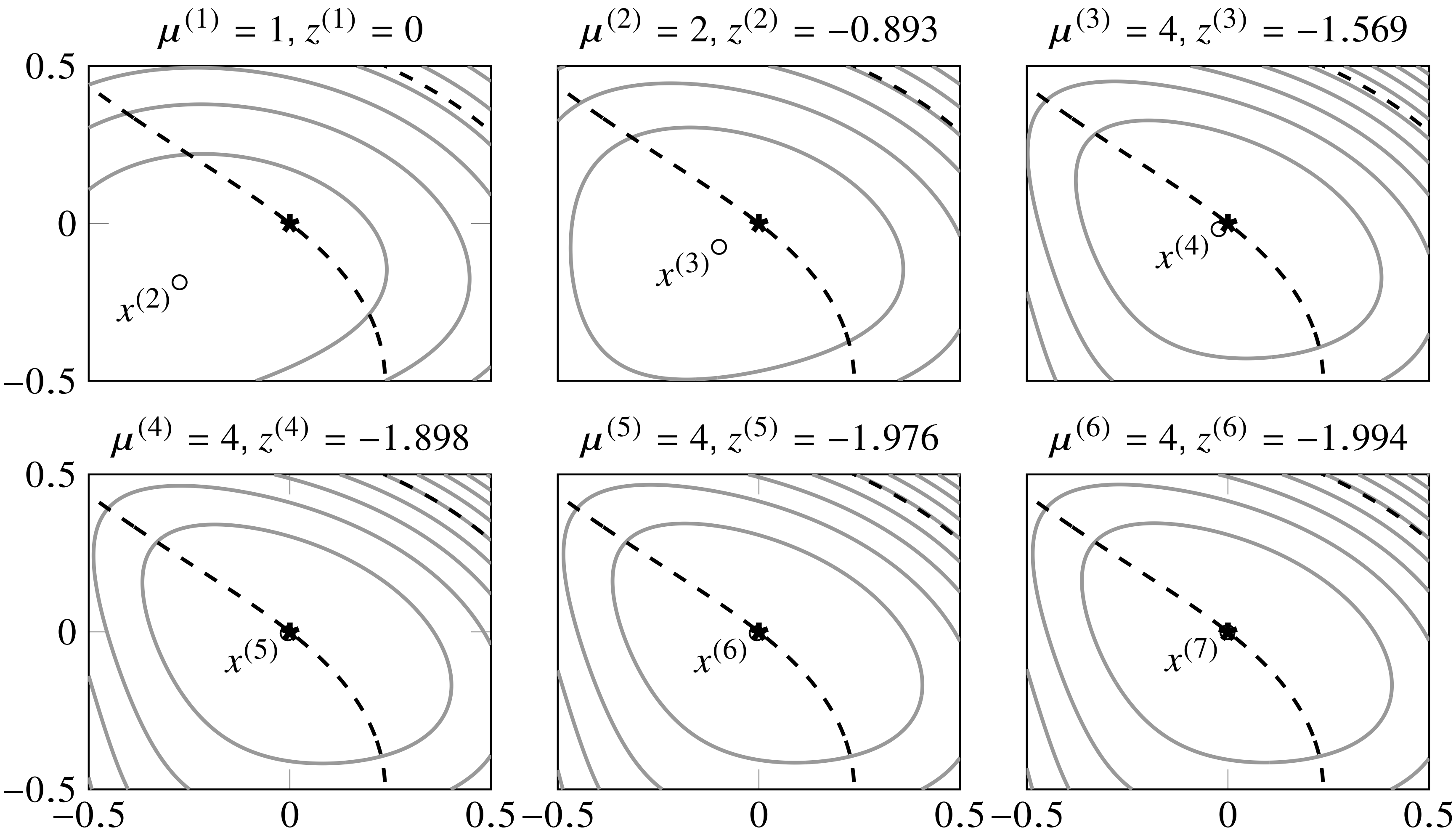
$$z^{(k+1)} = z^{(k)} + 2\mu^{(k)} g(x^{(k+1)}).$$

3. *penalty parameter update:*

$$\mu^{(k+1)} = \mu^{(k)} \quad \text{if } \|g(x^{(k+1)})\| < 0.25\|g(x^{(k)})\|, \qquad \mu^{(k+1)} = 2\mu^{(k)} \quad \text{otherwise}$$

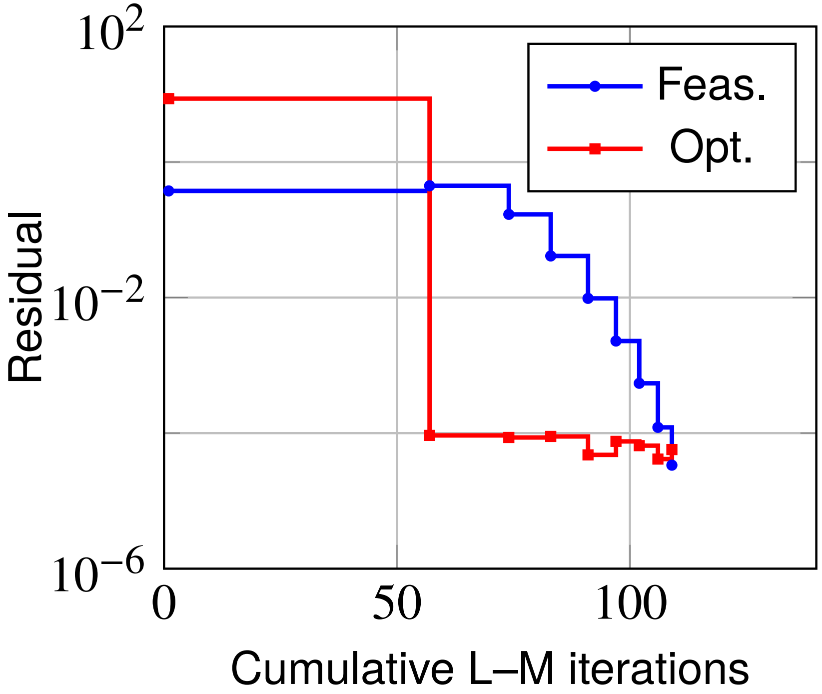
- iteration starts at $z^{(1)} = 0$, $\mu^{(1)} = 1$, some initial $x^{(1)}$
- μ is increased only when needed, more slowly than in penalty method
- continues until $g(x^{(k)})$ is sufficiently small (or iteration limit is reached)

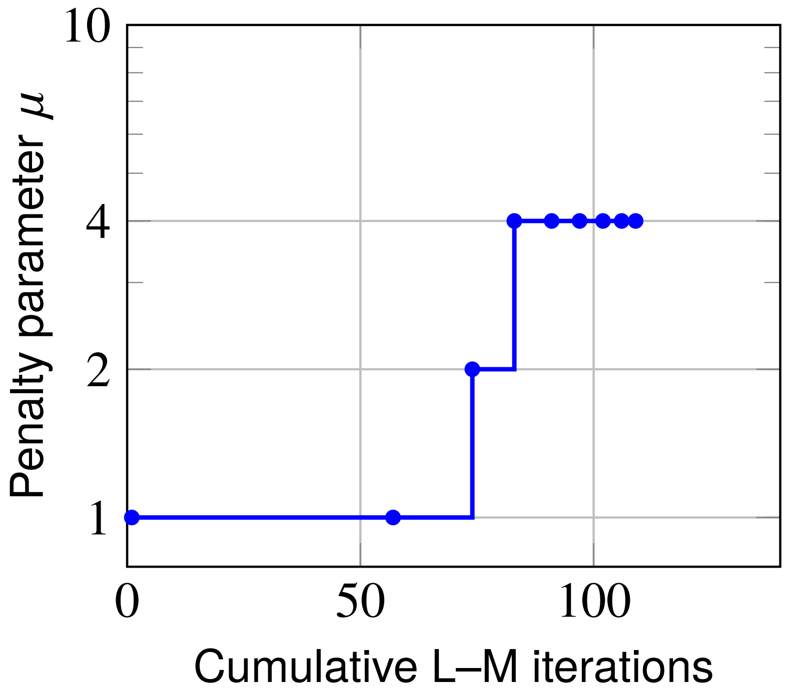
Example of slide 19.9



solid lines are contour lines of $L_{\mu^{(k)}}(x, z^{(k)})$

Convergence





- figure on the left shows residuals in optimality condition
- blue curve is norm of $g(x^{(k)})$
- red curve is norm of $2Df(x^{(k)})^T f(x^{(k)}) + Dg(x^{(k)})^T z^{(k)}$

Outline

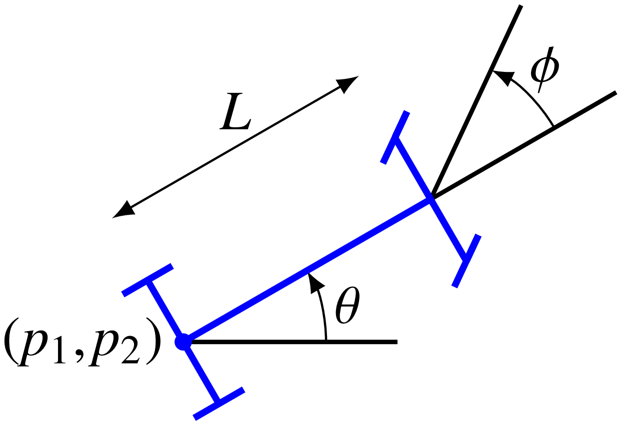
Constrained nonlinear least squares

Penalty method

Augmented Lagrangian method

Nonlinear control example

Simple model of a car



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

- $s(t)$ is speed of vehicle, $\phi(t)$ is steering angle
- $p(t)$ is position, $\theta(t)$ is orientation

Discretized model

- discretized model (for small time interval h):

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

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

$$f(x_k, u_k) = \left[\begin{array}{c} (x_k)_1 + h(u_k)_1\cos((x_k)_3) \\ (x_k)_2 + h(u_k)_1\sin((x_k)_3) \\ (x_k)_3 + h(u_k)_1\tan((u_k)_2)/L \end{array}\right]$$

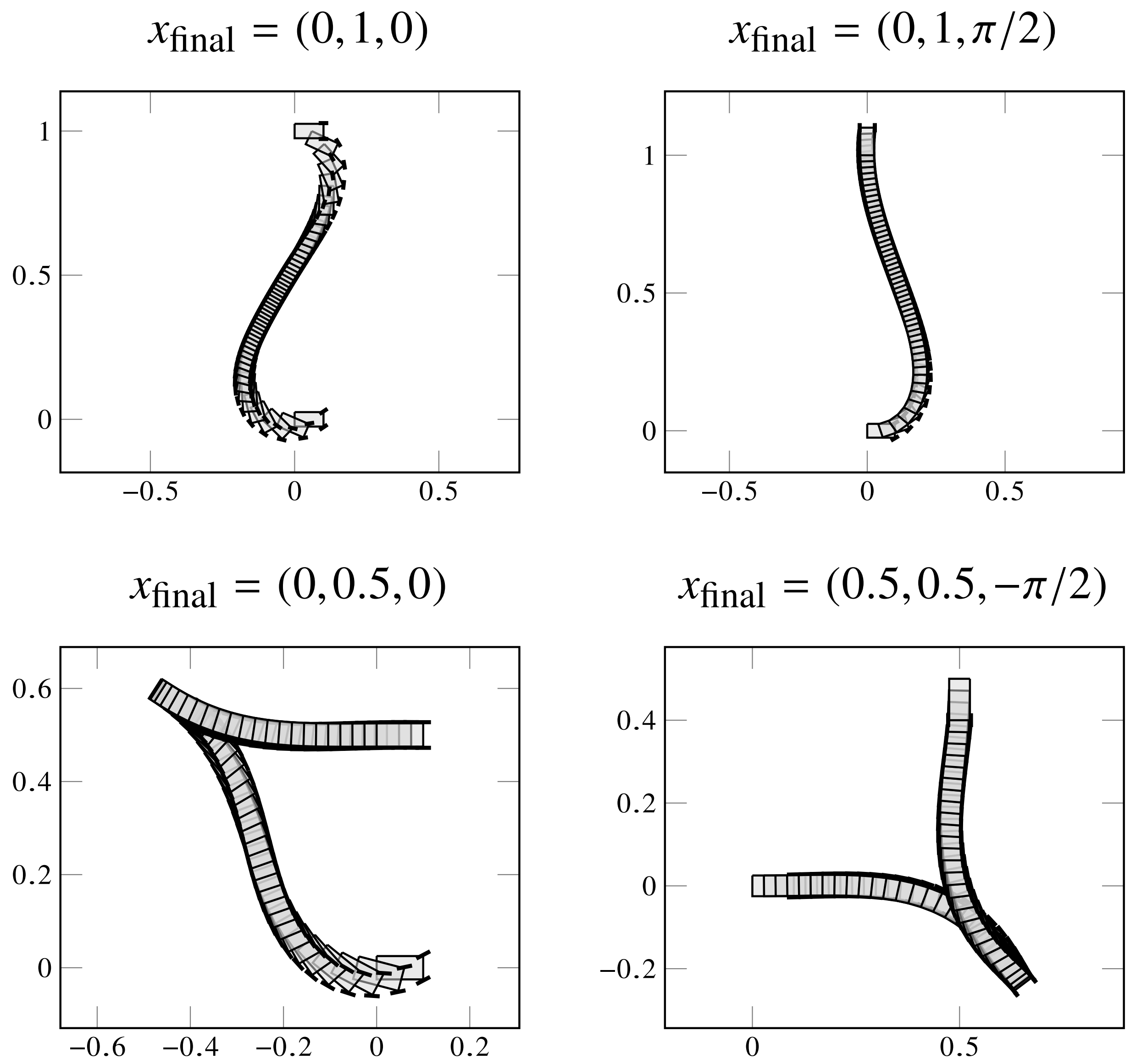
Control problem

- move car from given initial to desired final position and orientation
- using a small and slowly varying input sequence
- this is a constrained nonlinear least squares problem:

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

- variables are $u_1, \ldots, u_N$, $x_2, \ldots, x_N$

Four solution trajectories



Inputs for four trajectories