Stochastic optimization algorithms Lecture 3, 20200904

Classical optimization methods (ii)

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

CHALMERS

Recommended problems

- An exercise class will be held on 20200915. In that lecture, I will demonstrate the solution to several problems involving classical optimization , e.g. problems 2.12 and 2.13, as well as some problems related to EAs (Chapter 3).
- You should, as an exercise, solve the remaining problems in Chapter 2, i.e. Problems 2.1 – 2.11.
- Recommended problems: 2.1, 2.2, 2.3, 2.5, 2.8, 2.9, 2.11

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Today's learning goals

- After this lecture you should be able to
  - Describe and use Newton-Raphson's method
  - Define convex optimization problems
  - Describe and use the method of Lagrange multipliers
  - Describe and use an analytical method for constrained optimization
  - Describe and use the penalty method
  - List and discuss limitations of classical optimization

From last time...

- Iterative algorithms

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j + \eta_j \boldsymbol{d}_j$$

- Gradient descent

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j - \eta_j \nabla f(\boldsymbol{x}_j)$$

pp. 17-18

Newton-Raphson's method

- Consider the second-order expansion of $f(x)$:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)\ (x - x_0)^2 \equiv f_{[2]}(x)$$

- Minima occur at stationary points, where $f'_{[2]}(x) = 0$:

$$f'_{[2]}(x) = f'(x_0) + f''(x_0)(x - x_0) = 0 \Rightarrow x^* = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

pp. 21-24

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Newton-Raphson's method

- Thus, in general (iterative method!):

$$x_{j+1} = x_j - \frac{f'(x_j)}{f''(x_j)}$$

- Can be generalized to n dimensions (Newton's method), see p. 22.
- To do (for you!): Example 2.4.

pp. 21-24

Today’s learning goals

- After this lecture you should be able to
  - **Describe and use Newton-Raphson’s method**
  - Define convex optimization problems
  - Describe and use the method of Lagrange multipliers
  - Describe and use an analytical method for constrained optimization
  - Describe and use the penalty method
  - List and discuss limitations of classical optimization

Convex optimization problems

- A convex optimization problem is a special kind of constrained optimization problem, and it occurs if
  - f and g (the inequality constraints) are convex.
  - h (the equality constraints) are *affine*, i.e.

    $$h_i(\boldsymbol{x}) = {A_i}^T \boldsymbol{x} + b_i$$
- In that case S is convex and any local minimum is a global minimum.

pp. 24-25

Today's learning goals

- After this lecture you should be able to
  - Describe and use Newton-Raphson's method
  - **Define convex optimization problems**
  - Describe and use the method of Lagrange multipliers
  - Describe and use an analytical method for constrained optimization
  - Describe and use the penalty method
  - List and discuss limitations of classical optimization

Optimization under constraints

- Specific methods for solving convex optimization problems exist but will not be considered here.
- Instead, we will now consider three general methods for constrained optimization.
- The first two methods enumerate the possible optima, and are applicable when the number of variables is small, whereas the third method (the Penalty method) has more general applicability (and can also be used with stochastic optimization methods).

pp. 25-33

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

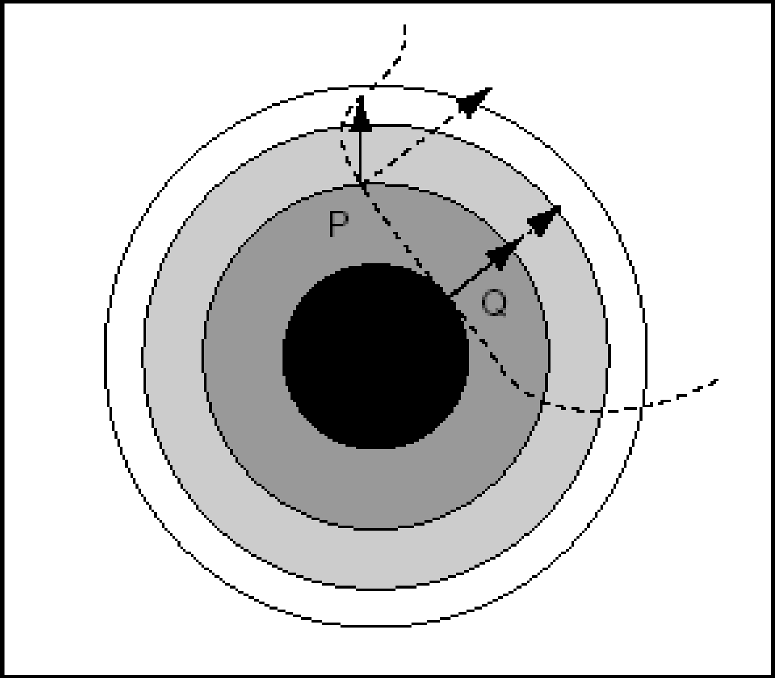
- Applicable in constrained optimization problems involving only equality constraints (so, m = 0, k > 0).
- The method is applicable for any dimension *n*, and for any number (k) of equality constraints.
- However, first, consider the case n = 2, k = 1, i.e. a two-dimensional problem with one equality constraint:

$$\text{minimize } f(x_1, x_2)$$
$$\text{subject to } h(x_1, x_2) = 0$$

pp. 25-28

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

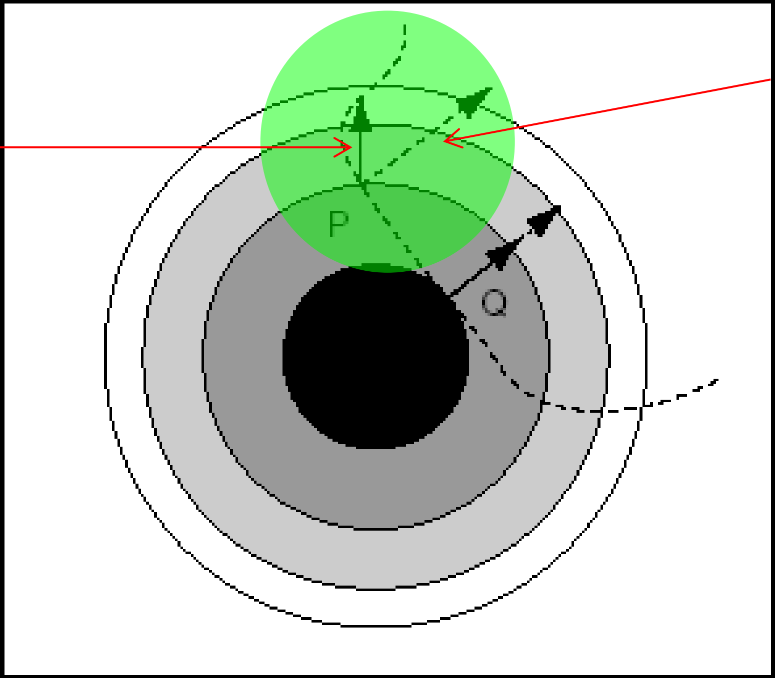


pp. 25-28

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

Gradient of f(x) (always orthogonal to the level curves)



Gradient of h(x) (always orthogonal to the constraint curve, i.e. $h(x_1,x_2)=0$).

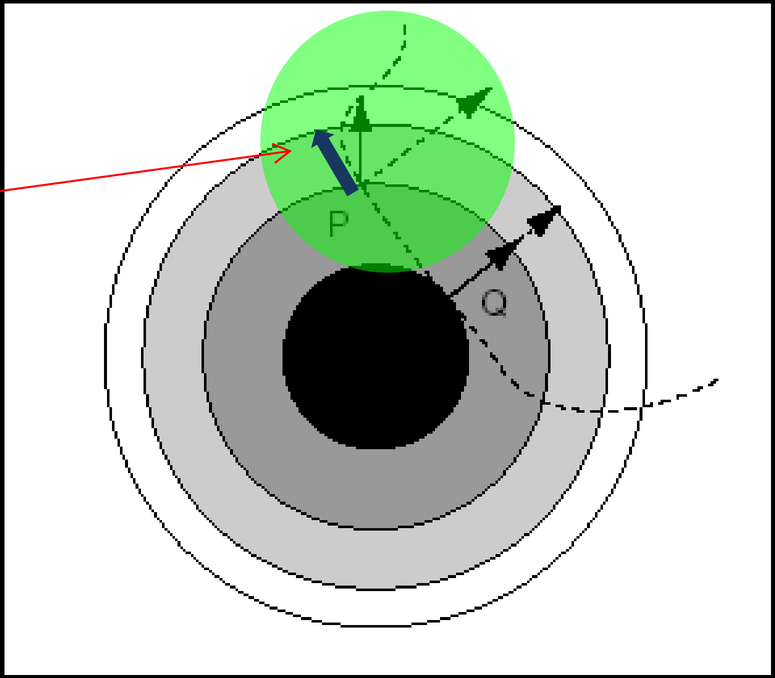
pp. 25-28

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

Towards larger values of f(x)



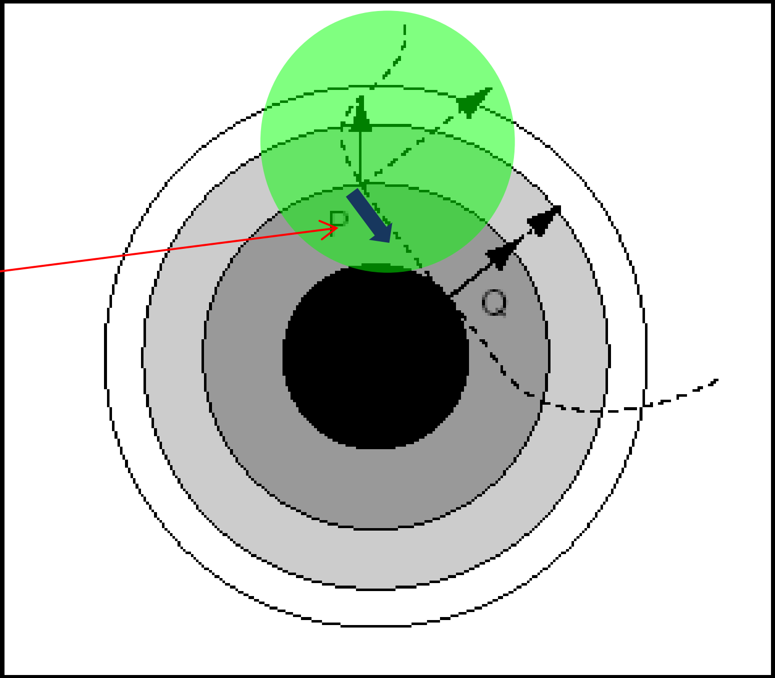
pp. 25-28

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

Towards smaller values of f(x)

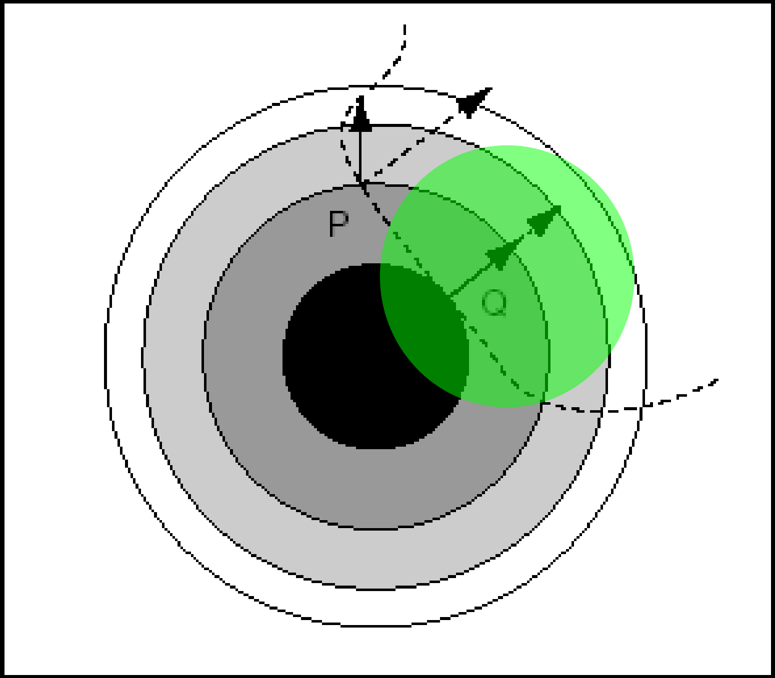


pp. 25-28

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:



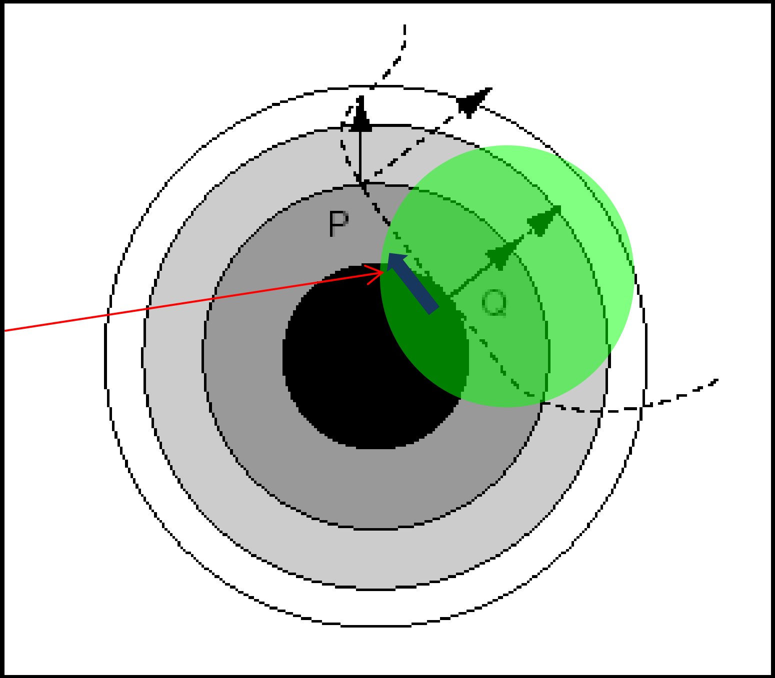
pp. 25-28

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

Towards larger values of f(x)

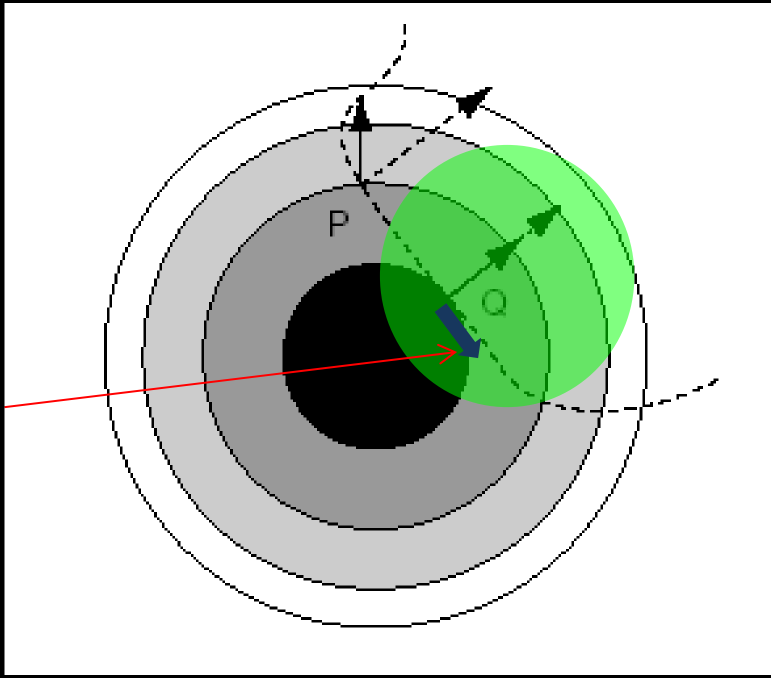


pp. 25-28

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

Towards larger values of f(x)

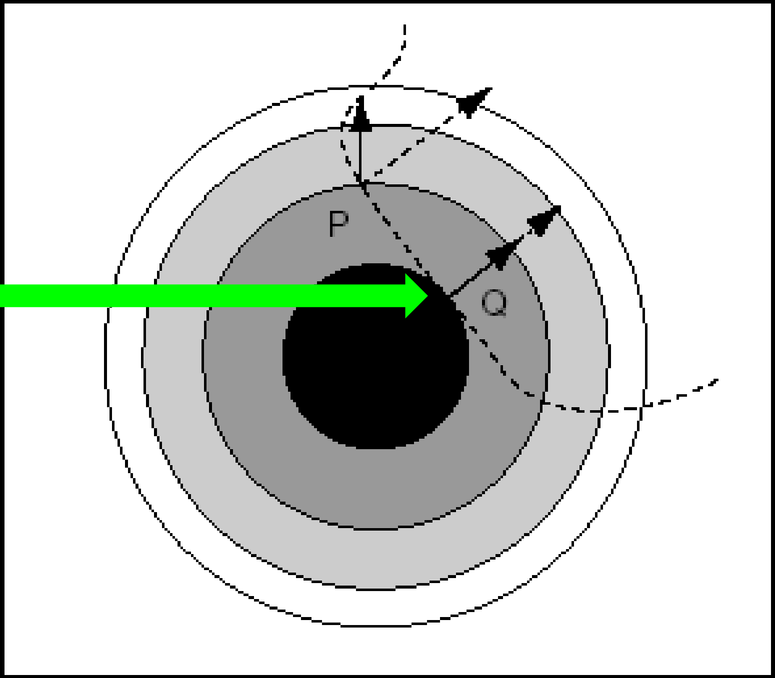


pp. 25-28

The method of Lagrange multipliers

- Movement along level curves (for a 2D problem – but can be generalized), in case of a single equality constraint:

Local optimum.
Here, $\nabla f = -\lambda \nabla h$



pp. 25-28

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

- Lagrange multiplier method. Consider the function

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2)$$

- Consider the equation $\nabla L = 0$:

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \lambda \frac{\partial h}{\partial x_1} = 0$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} + \lambda \frac{\partial h}{\partial x_2} = 0$$

$$\frac{\partial L}{\partial \lambda} = h = 0$$

pp. 25-28

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The method of Lagrange multipliers

- The local optima of f subject to the equality constraint(s) h_i = 0 can thus be found by computing the stationary points of L.
- Note: Finds both minima and maxima!
- To do (for you!): Example 2.6.

pp. 25-28

Today's learning goals

- After this lecture you should be able to
  - Describe and use Newton-Raphson's method
  - Define convex optimization problems
  - **Describe and use the method of Lagrange multipliers**
  - Describe and use an analytical method for constrained optimization
  - Describe and use the penalty method
  - List and discuss limitations of classical optimization

An analytical method for constr. opt.

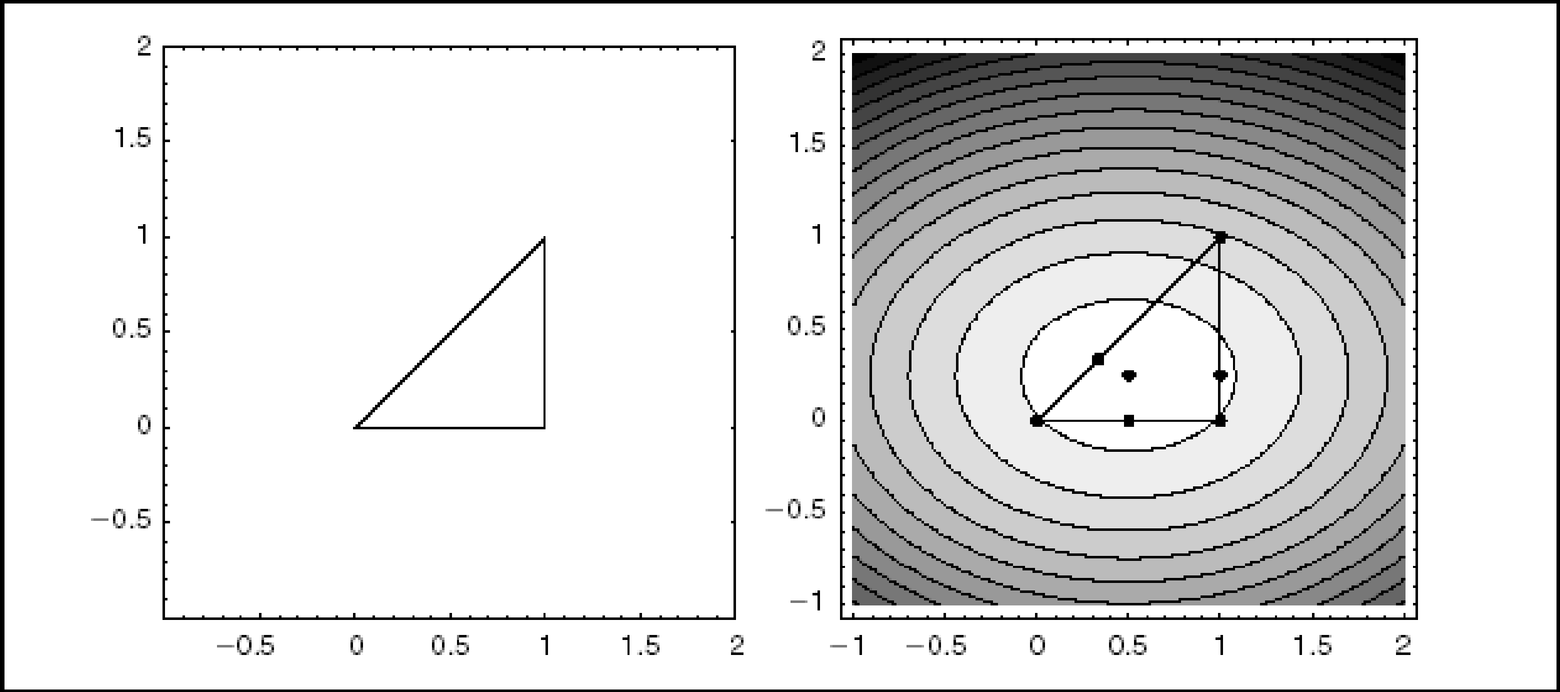
- Method description:
    1. Find the stationary points in the interior of S.
    2. Find the stationary points of the restriction of $f(x)$ to δS.
    3. Investigate the points one by one.
- This method can be used in low-dimensional problems.

pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

An analytical method for constr. opt.

- Example 2.7: $f(x_1, x_2) = x_1 - {x_1}^2 - 2{x_2}^2 + x_2$.



pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

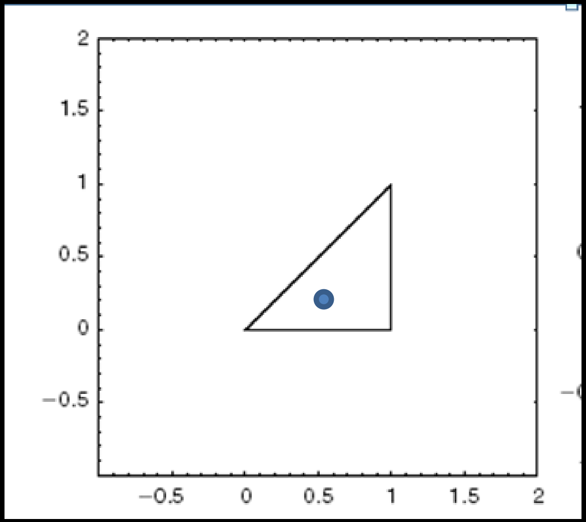
An analytical method for constr. opt.

- Step 1: Stationary points in the interior of S:

$$\frac{\partial f}{\partial x_1} = 1 - 2x_1 = 0$$

$$\frac{\partial f}{\partial x_2} = -4x_2 + 1 = 0$$

...from which one gets $x_1 = 1/2, x_2 = 1/4.$



- This point is clearly in the interior of S.
- Thus, the (only) candidate found in the interior of S is $P_1 = (\frac{1}{2}, \frac{1}{4})^T$

pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

An analytical method for constr. opt.

- Step 2: Stationary points on the boundary ∂S
- In this case, the boundary consists of three parts, which will be considered in sequence.
- Note that the corners will have to be treated separately.

pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

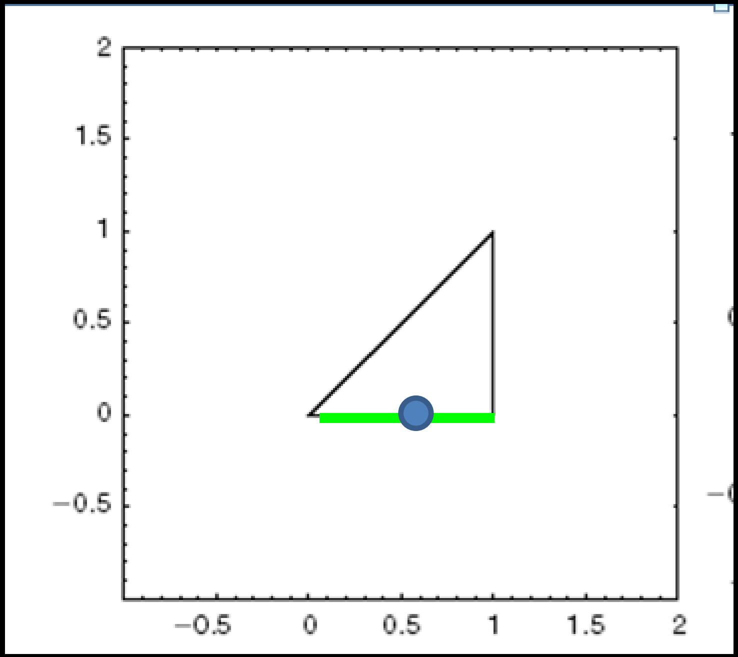
An analytical method for constr. opt.

- Step 2.1: $0 < x_1 < 1, x_2 = 0 \Rightarrow f(x_1, 0) = x_1 - {x_1}^2$
- Taking the derivative we get:

  $f'(x_{1,}0) = 1 - 2x_1$

  ... so that, solving $f'(x_{1,}0) = 0$,

  we obtain $x_1 = \frac{1}{2}$.
- Thus, this part of the boundary contributes the point $P_2 = (\frac{1}{2}, 0)^T$



pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

CHALMERS

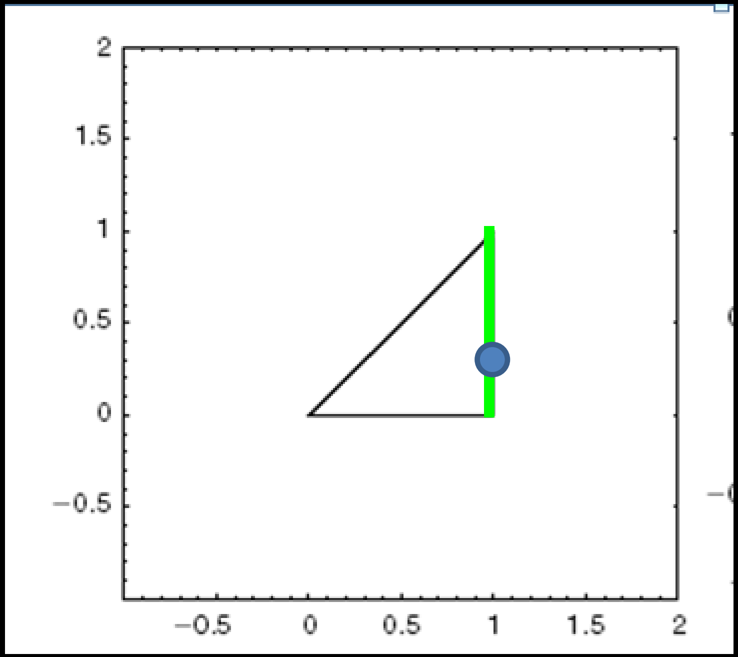
An analytical method for constr. opt.

- Step 2.2: $x_1 = 1, 0 < \ x_2 < 1, \Rightarrow f(1, x_2) = -2{x_2}^2 + x_2$
- Taking the derivative we get:

  $f'(1, x_2) = -4x_2 + 1$

  ... so that, solving $f'(1, x_2) = 0$,

  we obtain $x_2 = \frac{1}{4}$.



- Thus, this part of the boundary contributes the point $P_3 = (1, \frac{1}{4})^T$

pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

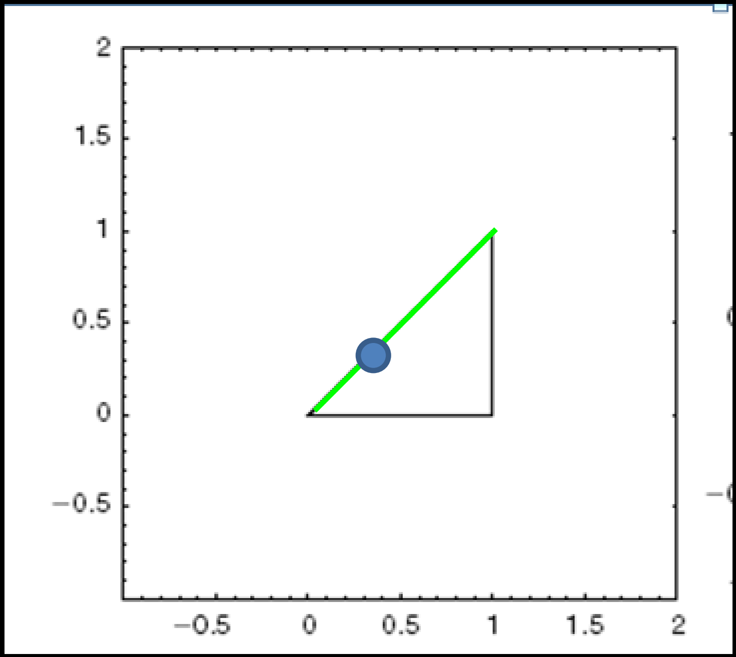
An analytical method for constr. opt.

- Step 2.3: $x_1 = x_2, 0 < x_1 < 1, \Rightarrow f(x_1, x_1) = 2x_1 - 3{x_1}^2$
- Taking the derivative we get:

  $f'(x_1, x_1) = 2 - 6x_1$

  ... so that, solving $f'(x_1, x_1) = 0$,

  we obtain $x_1 = \frac{1}{3}$.
- *Thus, this part of the boundary contributes* the point $P_4 = (\frac{1}{3}, \frac{1}{3})^T$

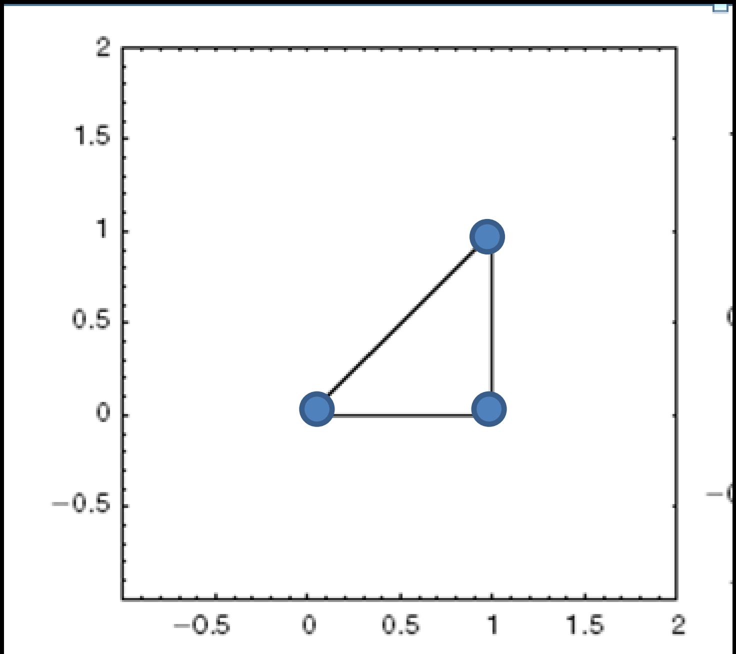


pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

An analytical method for constr. opt.

- Finally, the corners must be considered as well, contributing the points $P_5 = (0,0)^T$, $P_6 = (1,0)^T$, and $P_7 = (1,1)^T$
- Step 3: Examining these seven points one by one, one finds that the minimum occurs at one of the corner points, namely $P_7 = (1,1)^T$, where the function takes the value -1.



pp. 29-30

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Today’s learning goals

- After this lecture you should be able to
  - Describe and use Newton-Raphson’s method
  - Define convex optimization problems
  - Describe and use the method of Lagrange multipliers
  - **Describe and use an analytical method for constrained optimization**
  - Describe and use the penalty method
  - List and discuss limitations of classical optimization

The penalty method

- The penalty method transforms a constrained optimization problem to an unconstrained one.
- Consider the *penalty function* p:

$$p(\boldsymbol{x};\mu) = \mu\left(\sum_{i=1}^{m}(\max\{g_i(\boldsymbol{x}),0\})^2 + \sum_{i=1}^{k}(h_i(\boldsymbol{x}))^2\right)$$

pp. 30-33

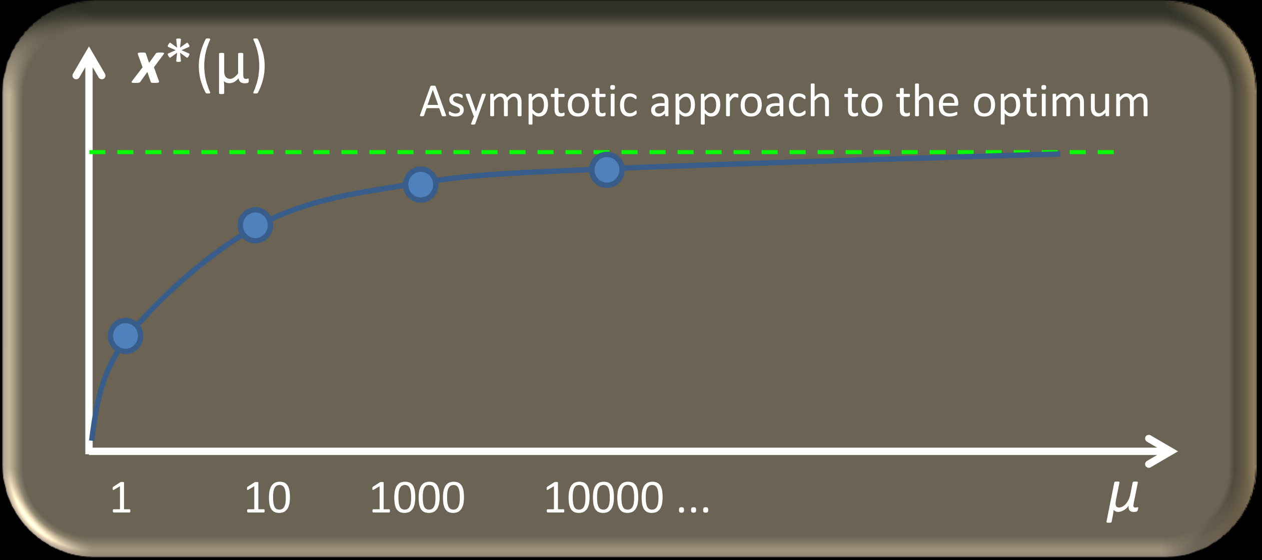
Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

The penalty method

- $p \geq 0$, with equality only if all constraints are satisfied.
- Thus, minimizing f subject to g and h is equivalent to minimizing $f_p(\boldsymbol{x};\mu) \equiv f(\boldsymbol{x}) + p(\boldsymbol{x};\mu)$ without constraints *as μ tends to infinity*.
- Numerical approach: Start with a small μ (e.g. 1), find the optimum (using e.g. gradient descent or any other method), then increase μ, again find the optimum etc. etc.

The penalty method

- As μ gets larger the optimum (normally) converges towards that of the constrained problem.



- To do (for you!): Examples 2.8 and 2.9.

Today's learning goals

- After this lecture you should be able to
  - Describe and use Newton-Raphson's method
  - Define convex optimization problems
  - Describe and use the method of Lagrange multipliers
  - Describe and use an analytical method for constrained optimization
  - **Describe and use the penalty method**
  - List and discuss limitations of classical optimization

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Limitations of classical optimization

- Classical methods are typically excellent *when they can be applied.*
- However, many kinds of (real-world) problems require other methods. Examples are problems where …
  - … the objective function cannot be specified explicitly as a mathematical function.
  - … the objective function is non-differentiable, and perhaps contains a mixture of (say) Boolean and numerical variables.
  - … the number of *variables* itself varies (for example in optimization of neural network of varying size).

Limitations of classical optimization

- In the rest of the course, we shall consider stochastic optimization methods, which can easily handle problems of the kinds just mentioned.
- **Important!** Make sure that you *attend the next few lectures* (as well as the programming session on Tuesday evening!).

Today's learning goals

- After this lecture you should be able to
  - Describe and use Newton-Raphson's method
  - Define convex optimization problems
  - Describe and use the method of Lagrange multipliers
  - Describe and use an analytical method for constrained optimization
  - Describe and use the penalty method
  - **List and discuss limitations of classical optimization**

Introductory programming problem

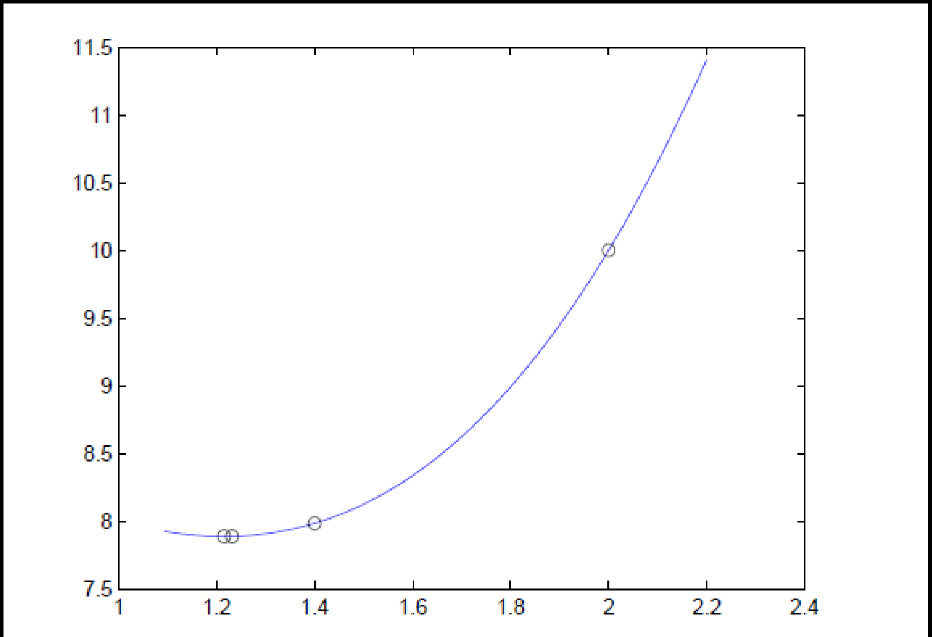
- Mandatory! Must be solved (and handed in, via Canvas) separately by each student.
- Available on the web page.
- Strict deadline (NOTE!) 20200911
- You will receive feedback (potentially relevant for your work with home problems 1 and 2) as soon as possible.

Introductory programming problem

- Important:
  - The main aim is for you to learn how to write clear, well-structured program code (in Matlab, in this case).
  - Make sure to *follow the coding standard* (available on the course web page!).
  - Before submitting the solution (via Canvas), check that it can run on the Matlab version available at Chalmers.

Introductory programming problem

- Problem: Implement a Newton-Raphson solver for polynomials, using Matlab.



Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde