

Methods for non-linear problems

Line search

Niclas Börlin

5DA001 Non-linear Optimization

1 / 32

The Newton Method

Summary

- ▶ Naive Newton (1-D): Use Newton-Raphson approach on first-order conditions

$$f'(x) = 0.$$

- ▶ This gives the iteration sequence

$$x_{k+1} = x_k + \frac{-f'(x_k)}{f''(x_k)}.$$

- ▶ In n -D, the first-order conditions are

$$\nabla f(x) = 0.$$

- ▶ The iteration (Newton) sequence becomes

$$x_{k+1} = x_k + \underbrace{\nabla^2 f(x_k)^{-1}(-\nabla f(x_k))}_{p_k}.$$

2 / 32

The Newton Method

Summary

- ▶ The identity

$$p_k = \nabla^2 f(x_k)^{-1}(-\nabla f(x_k))$$

is equivalent to that p_k is the solution to the Newton Equation:

$$\boxed{\nabla^2 f(x_k)p_k = -\nabla f(x_k)}.$$

3 / 32

The Newton Method

Summary

- ▶ Problem: The Newton method converges to a **stationary point**, not necessarily a minimizer.
- ▶ How to make the Newton method converge to a **minimizer**?

4 / 32

The Newton Method

Summary

- Answer (part 1): Consider the quadratic polynomial

$$m_k(x_k + p) \equiv f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p.$$

- The first order condition on m_k is

$$\nabla_p m_k(x_k + p) = \nabla f(x_k) + \nabla^2 f(x_k) p = 0.$$

- The solution is

$$p = \nabla^2 f(x_k)^{-1} (-\nabla f(x_k)).$$

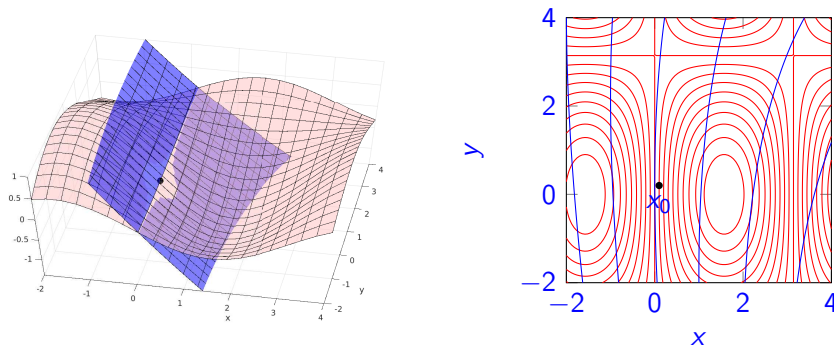
- If the Hessian $\nabla^2 f(x_k)$ is positive definite, the update p is the minimizer for m_k .
- If all $\nabla^2 f(x_k)$ are positive definite, the Newton sequence will converge towards a minimizer.

5 / 32

2D function — $f(x) = -\sin x_1 \cos x_2 / 2$

$$x_k = \begin{pmatrix} 0.10 \\ 0.20 \end{pmatrix}, \nabla^2 f(x_k) = \begin{pmatrix} 0.10 & 0.05 \\ 0.05 & 0.02 \end{pmatrix}, \lambda = \begin{pmatrix} 0 \\ 0.12 \end{pmatrix},$$

$$p = \begin{pmatrix} \infty \\ -\infty \end{pmatrix}, x_{k+1} = \begin{pmatrix} \infty \\ -\infty \end{pmatrix}.$$



7 / 32

The Newton Method

Summary

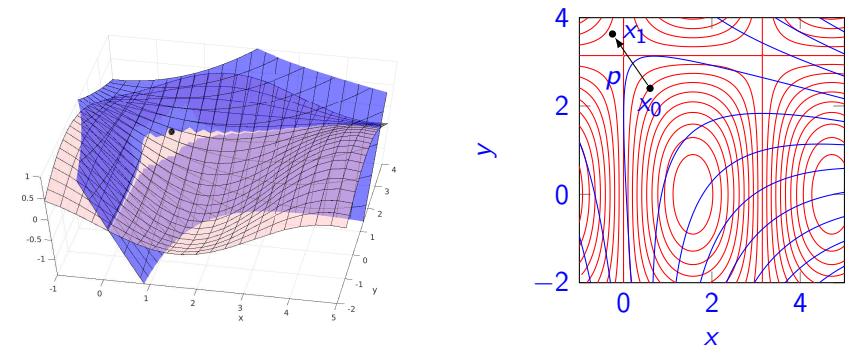
- Reminder: A positive definite matrix has positive eigenvalues.
- If the Hessian $\nabla^2 f(x_k)$ is not positive definite, we have three cases:
 - $\nabla^2 f(x_k)$ has a zero eigenvalue: $\nabla^2 f(x_k)^{-1}$ does not exist and Naive Newton will fail.
 - $\nabla^2 f(x_k)$ has non-zero eigenvalues of mixed signs: p is a saddle point on m_k .
 - $\nabla^2 f(x_k)$ has all negative eigenvalues: p is a maximizer on m_k .
- In either case, Naive Newton is not guaranteed to converge towards a minimizer. (It probably won't.)

6 / 32

2D function — $f(x) = -\sin x_1 \cos x_2 / 2$

$$x_k = \begin{pmatrix} 0.60 \\ 2.40 \end{pmatrix}, \nabla^2 f(x_k) = \begin{pmatrix} 0.20 & 0.38 \\ 0.38 & 0.05 \end{pmatrix}, \lambda = \begin{pmatrix} -0.26 \\ 0.52 \end{pmatrix},$$

$$p = \begin{pmatrix} -0.85 \\ 1.23 \end{pmatrix}, x_{k+1} = \begin{pmatrix} -0.25 \\ 3.63 \end{pmatrix}.$$

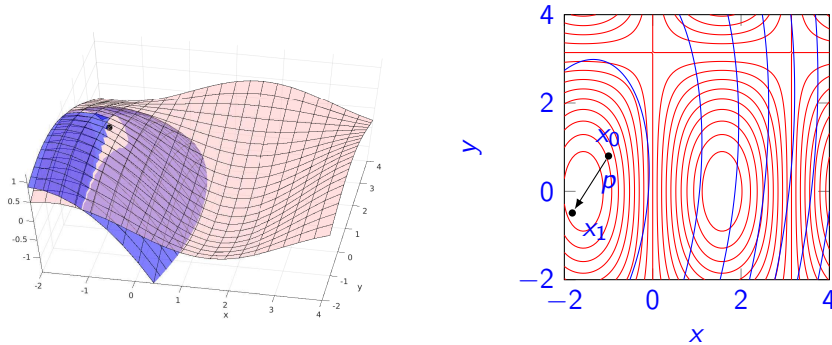


8 / 32

2D function — $f(x) = -\sin x_1 \cos x_2/2$

$$x_k = \begin{pmatrix} -1.00 \\ 0.80 \end{pmatrix}, \nabla^2 f(x_k) = \begin{pmatrix} -0.78 & 0.11 \\ 0.11 & -0.19 \end{pmatrix}, \lambda = \begin{pmatrix} -0.79 \\ -0.18 \end{pmatrix},$$

$$p = \begin{pmatrix} -0.82 \\ -1.29 \end{pmatrix}, x_{k+1} = \begin{pmatrix} -1.82 \\ -0.49 \end{pmatrix}.$$



9 / 32

How to make the Newton method globally convergent

- Answer (part 2): Make the Newton method a line search method.
 - Rewrite update formula

$$x_{k+1} = x_k + \alpha_k p_k.$$

- First compute the search direction p_k , then compute the step length $\alpha_k > 0$.
- Note: $\alpha_k = 1$ gives Naive Newton.
- Require p_k and α_k to satisfy:
 - 2a) Each search direction p_k produces “sufficient descent”.
 - 2b) Each search direction p_k is “gradient related”.
 - 2c) Each step length α_k produces “sufficient decrease”.
 - 2d) Each step length α_k is not “too small”.

10 / 32

How to make the Newton method globally convergent

- 2a+2b satisfied if the Hessian $H_k = \nabla^2 f(x_k)$ positive definite and its eigenvalues bounded from above and below:

$$0 < \ell_{\min} \leq \lambda(\nabla^2 f(x_k)) \leq \ell_{\max} < \infty.$$

- Solution: Modify the LDL^T -factorization used in the solution of the Newton equation to have $d_{ii} \geq d_{\min} > 0$.
- The modified $H'_k = LD'L^T$ will be have eigenvalues bounded from below away from 0.
- A search direction computed from

$$H'_k p = -\nabla f(x_k)$$

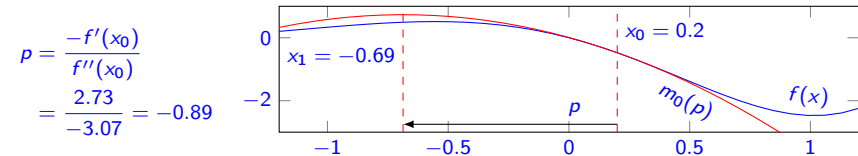
will be a descent direction and satisfy 2a+2b.

11 / 32

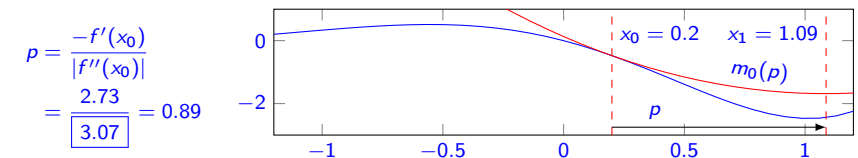
The Newton direction and descent

Modifying the Hessian

- Without modification:



- With modification (1):



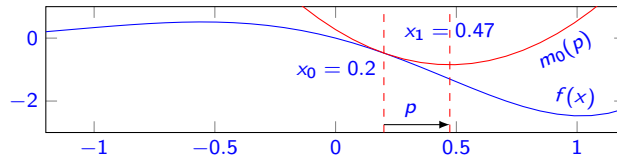
12 / 32

The Newton direction and descent

Modifying the Hessian

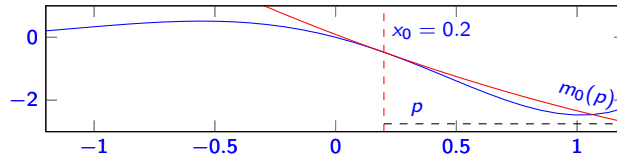
► With modification (2):

$$p = \frac{-f'(x_0)}{10} = \frac{2.73}{10} = 0.27$$



► With modification (3):

$$p = \frac{-f'(x_0)}{1} = 2.73$$



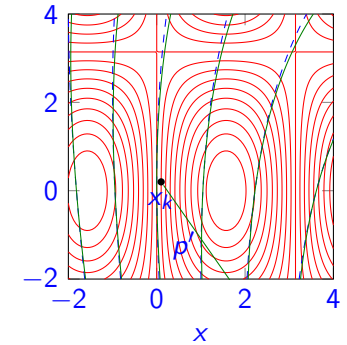
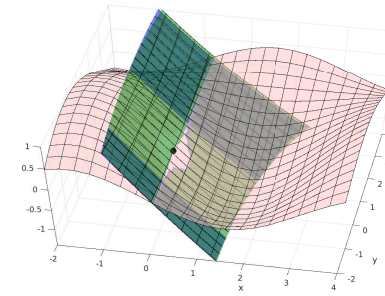
13 / 32

2D function — $f(x) = -\sin x_1 \cos x_2/2$

$$x_k = \begin{pmatrix} 0.10 \\ 0.20 \end{pmatrix}, H_k = \begin{pmatrix} 0.10 & 0.05 \\ 0.05 & 0.02 \end{pmatrix}, \lambda = \begin{pmatrix} 0 \\ 0.12 \end{pmatrix}, D = \begin{pmatrix} 0.10 & 0 \\ 0 & 0.00 \end{pmatrix}, L = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix},$$

$$D' = \begin{pmatrix} 0.10 & 0 \\ 0 & 0.01 \end{pmatrix}, H'_k = \begin{pmatrix} 0.10 & 0.05 \\ 0.05 & 0.03 \end{pmatrix}, \lambda' = \begin{pmatrix} 0.01 \\ 0.13 \end{pmatrix},$$

$$p = \begin{pmatrix} \infty \\ -\infty \end{pmatrix}, x_{k+1} = \begin{pmatrix} \infty \\ -\infty \end{pmatrix}, p' = \begin{pmatrix} 34.97 \\ -50.00 \end{pmatrix}, x'_{k+1} = \begin{pmatrix} 35.07 \\ -49.80 \end{pmatrix}.$$



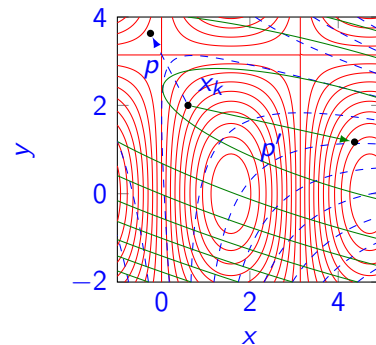
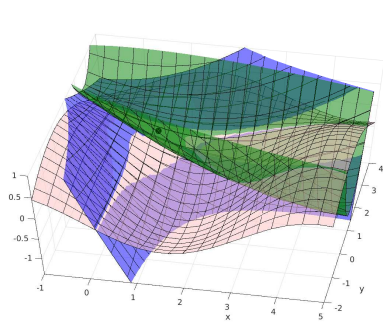
14 / 32

2D function — $f(x) = -\sin x_1 \cos x_2/2$

$$x_k = \begin{pmatrix} 0.60 \\ 2.40 \end{pmatrix}, H_k = \begin{pmatrix} 0.20 & 0.38 \\ 0.38 & 0.05 \end{pmatrix}, \lambda = \begin{pmatrix} -0.26 \\ 0.52 \end{pmatrix}, D = \begin{pmatrix} 0.20 & 0 \\ 0 & -0.67 \end{pmatrix}, L = \begin{pmatrix} 1 & 0 \\ 1.9 & 1 \end{pmatrix},$$

$$D' = \begin{pmatrix} 0.20 & 0 \\ 0 & 0.67 \end{pmatrix}, H'_k = \begin{pmatrix} 0.20 & 0.38 \\ 0.38 & 1.39 \end{pmatrix}, \lambda' = \begin{pmatrix} 0.09 \\ 1.51 \end{pmatrix},$$

$$p = \begin{pmatrix} -0.85 \\ 1.23 \end{pmatrix}, x_{k+1} = \begin{pmatrix} -0.25 \\ 3.63 \end{pmatrix}, p' = \begin{pmatrix} 3.77 \\ -1.23 \end{pmatrix}, x'_{k+1} = \begin{pmatrix} 4.37 \\ 1.17 \end{pmatrix}.$$



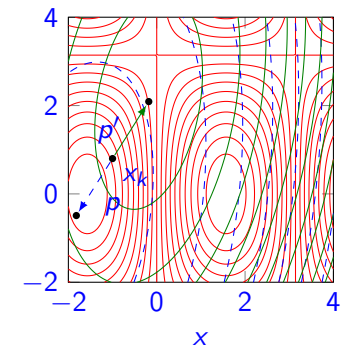
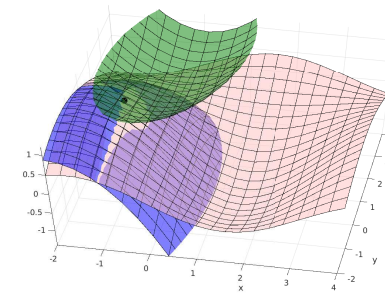
15 / 32

2D function — $f(x) = -\sin x_1 \cos x_2/2$

$$x_k = \begin{pmatrix} -1.00 \\ 0.80 \end{pmatrix}, H_k = \begin{pmatrix} -0.78 & 0.11 \\ 0.11 & -0.19 \end{pmatrix}, \lambda = \begin{pmatrix} -0.79 \\ -0.18 \end{pmatrix}, D = \begin{pmatrix} -0.78 & 0 \\ 0 & -0.18 \end{pmatrix}, L = \begin{pmatrix} 1 & 0 \\ -0.1 & 1 \end{pmatrix},$$

$$D' = \begin{pmatrix} 0.78 & 0 \\ 0 & 0.18 \end{pmatrix}, H'_k = \begin{pmatrix} 0.78 & -0.11 \\ -0.11 & 0.19 \end{pmatrix}, \lambda' = \begin{pmatrix} 0.18 \\ 0.79 \end{pmatrix},$$

$$p = \begin{pmatrix} -0.82 \\ -1.29 \end{pmatrix}, x_{k+1} = \begin{pmatrix} -1.82 \\ -0.49 \end{pmatrix}, p' = \begin{pmatrix} 0.82 \\ 1.29 \end{pmatrix}, x'_{k+1} = \begin{pmatrix} -0.18 \\ 2.09 \end{pmatrix}.$$



16 / 32

Line search

Line search

- ▶ In a line search algorithm, the algorithm chooses a **search direction** p_k and tries to solve the following one-dimensional minimization problem

$$\min_{\alpha > 0} f(x_k + \alpha p_k),$$

where the scalar α is called the **step length**.

17 / 32

18 / 32

Line search

- ▶ Each iteration of a line search method computes a search direction p_k and then decides how far to move along that direction.
- ▶ The next iteration is given by

$$x_{k+1} = x_k + \alpha_k p_k$$

- ▶ We will require p_k to be a descent direction. This assures that the objective function will decrease

$$f(x_k + \alpha_k p_k) < f(x_k)$$

for some small $\alpha_k > 0$.

19 / 32

Exact and inexact line searches

- ▶ Consider the function

$$\phi(\alpha) = f(x_k + \alpha p_k), \quad \alpha > 0.$$

- ▶ Ideally we would like to find the global minimizer of ϕ for every iteration. This is called an **exact** line search.
- ▶ However, it is possible to construct **inexact** line search methods that produce an adequate reduction of f at a minimal cost.
- ▶ Inexact line search methods construct a number of candidate values for α and stop when certain conditions are satisfied.

20 / 32

The Sufficient Decrease Condition

- ▶ Mathematically, the descent condition $f(x_k + \alpha p_k) < f(x_k)$ is not enough to guarantee convergence.
- ▶ Instead, the **sufficient decrease** condition is formulated from the linear Taylor approximation of $\phi(\alpha)$

$$\phi(\alpha) \approx \phi(0) + \alpha \phi'(0)$$

or

$$f(x_k + \alpha p_k) \approx f(x_k) + \alpha \nabla f_k^T p_k.$$

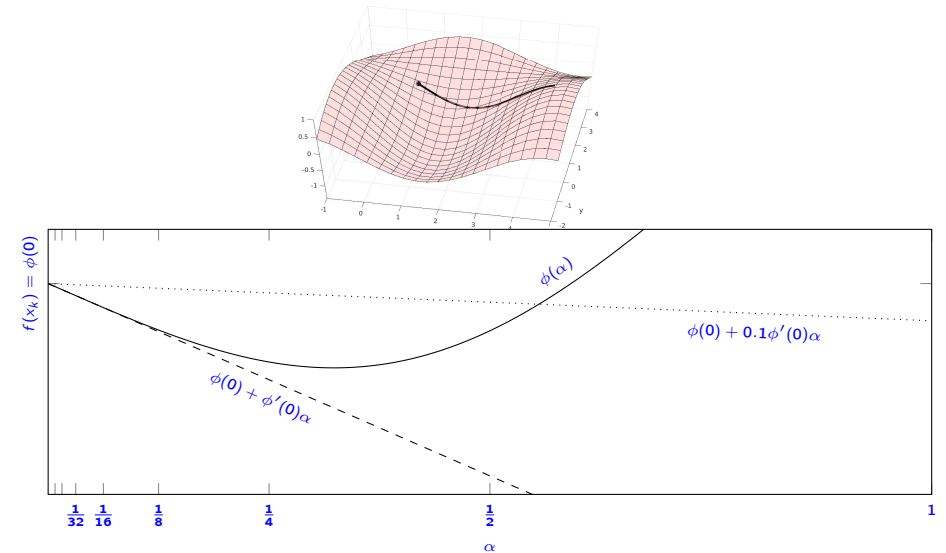
- ▶ The sufficient decrease condition states that the new point must at least produce a fraction $0 < c_1 < 1$ of the decrease predicted by the Taylor approximation, i.e.

$$f(x_k + \alpha p_k) < f(x_k) + c_1 \alpha \nabla f_k^T p_k.$$

- ▶ This condition is sometimes called the **Armijo** condition.

21 / 32

The Armijo Condition



22 / 32

Backtracking

- ▶ The sufficient decrease condition alone is not enough to guarantee convergence, since it is satisfied for arbitrarily small values of α .
- ▶ The sufficient decrease condition has to be combined with a strategy that favours large step lengths over small.
- ▶ A simple such strategy is called **backtracking**: Accept the first element of the sequence

$$1, \frac{1}{2}, \frac{1}{4}, \dots, 2^{-i}, \dots$$

that satisfies the sufficient decrease condition. Such a step length always exist.

- ▶ Large step lengths are tested before small ones. Thus, the step length will not be too small.
- ▶ This technique works well for Newton-type algorithms.

23 / 32

The Curvature Condition

- ▶ Another approximation to the solution of

$$\min_{\alpha > 0} \phi(\alpha) \equiv f(x_k + \alpha p_k)$$

is to solve for $\phi'(\alpha) = 0$, which is approximated to the condition

$$|\phi'(\alpha_k)| \leq c_2 |\phi'(0)|,$$

where c_2 is a constant $c_1 < c_2 < 1$.

- ▶ Since $\phi'(\alpha) = p_k^T \nabla f(x_k + \alpha p_k)$, we get

$$|p_k^T \nabla f(x_k + \alpha_k p_k)| \leq c_2 |p_k^T \nabla f(x_k)|.$$

This condition is called the **curvature condition**.

24 / 32

The Wolfe Condition

- The sufficient decrease condition and the curvature condition

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k,$$

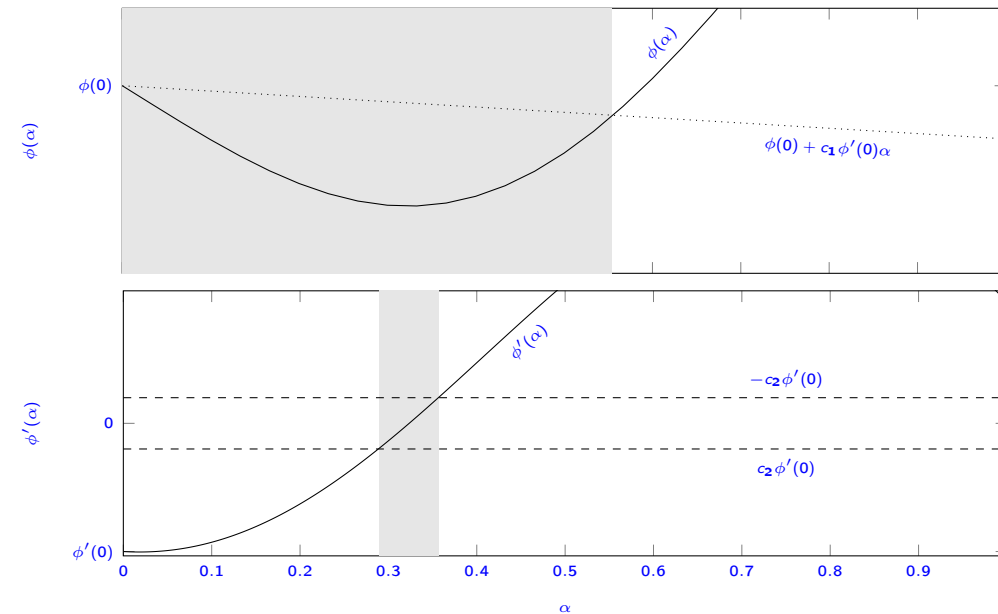
$$|p_k^T \nabla f(x_k + \alpha_k p_k)| \leq c_2 |p_k^T \nabla f(x_k)|,$$

where $0 < c_1 < c_2 < 1$, are collectively called the **strong Wolfe conditions**.

- Step length methods that use the Wolfe conditions are more complicated than backtracking.
- Several popular implementations of nonlinear optimization routines are based on the Wolfe conditions, notably the BFGS quasi-Newton method.

25 / 32

The Wolfe Condition



26 / 32

The modified Newton algorithm with line search

- Specify a starting approximation x_0 and a convergence tolerance ε .
- Repeat for $k = 0, 1, \dots$
 - If $\|\nabla f(x_k)\| < \varepsilon$, stop.
 - Compute the modified **LDL^T** factorization of the Hessian.
 - Solve

$$(LDL^T)p_k^N = -\nabla f(x_k)$$

for the search direction p_k^N .

- Perform an Armijo line search with backtracking to determine the new approximation $x_{k+1} = x_k + \alpha_k p_k^N$.

27 / 32

Globalization strategies

- The line search approach is sometimes called a **globalization strategy**, since it modifies a “core” method (typically locally convergent) to become globally convergent.
- There are two efficiency requirements on any globalization strategy:
 - Far from the solution, they should **stop** the methods from **going out of control**.
 - Close to the solution, when the “core” method is efficient, they should **interfere as little as possible**.

28 / 32

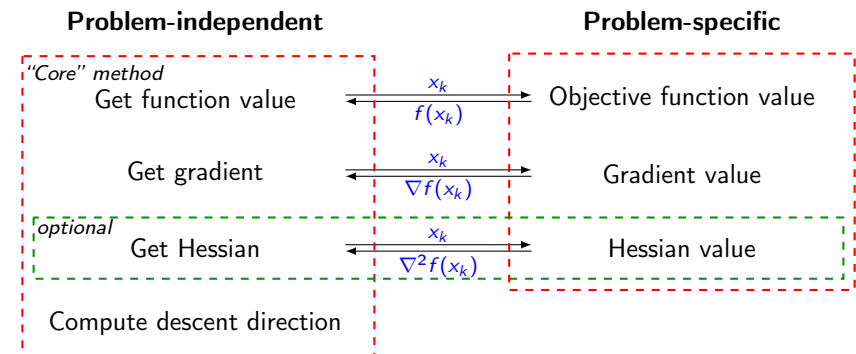
Globalization strategies

- ▶ The modified Newton method satisfy both efficiency requirements:
- ▶ Far from the solution, the modified LDL^T method will ensure that all search directions p_k are descent directions, even if the Hessian is not positive definite.
- ▶ Furthermore, if the p_k is long and does not produce a good reduction of $f(x_k + \alpha p_k)$, the line search strategy will try successively shorter steps, until a good enough point is found.
- ▶ Close to the solution, the Hessian is usually positive definite and the LDL^T method will not interfere.
- ▶ The backtracking algorithm always tries “full” steps ($\alpha = 1$) first. Near the solution, the Taylor polynomial will usually be a good approximation and $\alpha = 1$ will be accepted.

29 / 32

Abstraction level

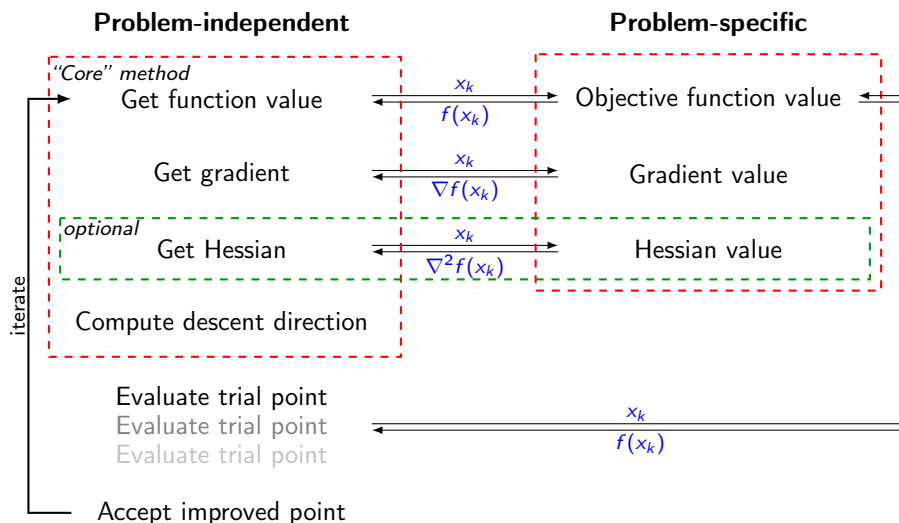
Core method



30 / 32

Abstraction level

Globalization strategy



31 / 32

Questions

- ▶ For a well-behaved problem, i.e. $\nabla^2 f(x_*)$ positive definite, how will the step lengths behave near the solution (small/large)?
- ▶ Does line search have a memory? (I.e. does the step length α_k at one iteration affect the step length at the next iteration?)
- ▶ Do you think it is worthwhile to try to solve the linesearch problem as accurately as possible, or just pick a quick solution?

32 / 32