OCS Hints for Questions

Julian Wolf

January 27, 2017

Derivations

Die Antworten sind teilweise unvollständig, einerseits, weil er die Antworten als "Eh Klar" abgestempelt hat, anderer seits weil er so schnell durchging, dass ein Mitschreiben nicht mehr möglich war.

1. Draw level lines and arrows
   - objective function is the function we want to minimize
   - constraint set is a set of functions
   - optimal solution: find $f(x^*) \leq f(x), \forall x \in X$
   - level set: compareable to level lines of terrain, convex function => convex level set (but there are non convex fct with convex level sets),
2. - **Linear**: Objective Function and Constraints may only be linear $min\ c^T x, s.t.\ Ax \leq b, x \geq 0$
     Polynomial solvable
   - **Non Linear**: Objective Function and Constriants may be non linear $min\ \frac{1}{2}x^T Qx + c^T x, s.t.\ Ax \leq b, Ex = d$
     Q symmetrical and pos. definite, polynomial solvable
   - **Quadratic**: objective function is quadratic, constraints are linear $min_{x \in \mathbb{R}}\ f_0(x)$ `(objective)`,
     $s.t.\ f_i(x) \leq i = 0..m$ `(contraints)`
     polynomial time
   - **convex set**: $\alpha x + (1-\alpha)y\ in X, \forall x, y \in X, \alpha \in [0,1]$
   - **convex fct**: $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall x, y \in X, \alpha \in [0,1]$
3. - When hessian is strictly positive, it is a strict global maximum
   - **unconst Local minimum:** $f(x^*) \leq f(x), \forall x$ `with` $||x - x^*|| \leq \varepsilon$
   - **unconst global minimum:** $f(x^*) \leq f(x), \forall x \in \mathbb{R}$

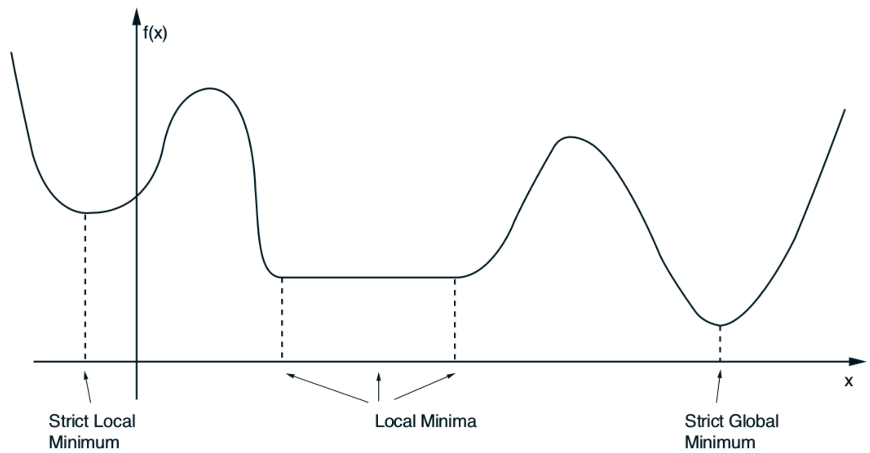

Figure 1: Local/Global minimas

4.
- If positive and negative Eigenvalues, we can not define convexity
- First order necessessary optimality condition: $\nabla f(x^*) = 0$
- Second order necessessary optimality condition: $\nabla^2 f(x^*)$ is positive semi definite

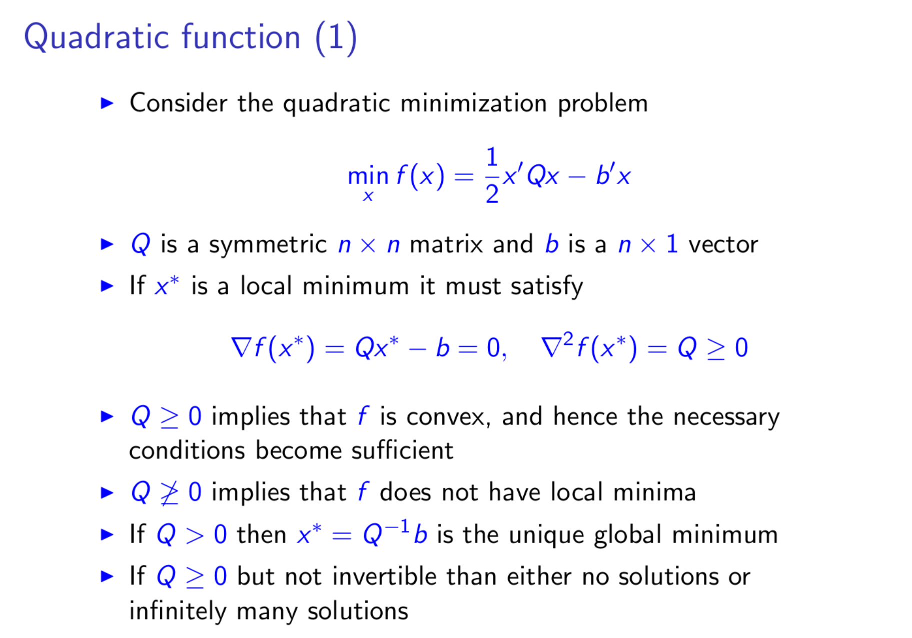

Figure 2: Different scenarios for Q

5.
- Descent direction: angle of step and derivation direction $< 90°$
- General form of gradient method:
  1. Choose an initial vector $x^0 \in \mathbb{R}^n$
  2. Choose a descent direction d^k that satisfies $\nabla f(x^k)'d^k < 0$
  3. Choose a positive step size α^k
  4. Compute the new vector as
  $x^{k+1} = x^k + \alpha^k d^k$
  5. Set $k = k + 1$ and goto 2

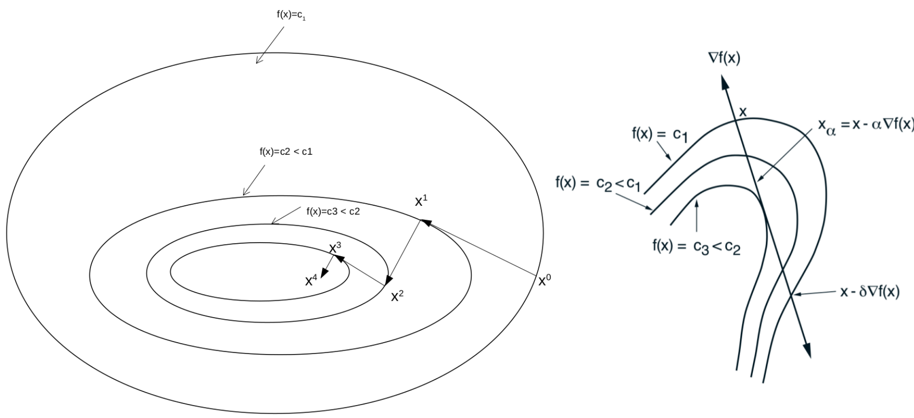

Figure 3: Simple descent direction

6. $d^k = -D^k \nabla f(x^k)$

    - Identity: $D^k = I$, = Gradient descent, zig zagging problem, very bad on Rosenbrock Fct
    - Hessian: $D^k = \nabla^2 f(x^k)$, = Newtons method, very fast convergence, very good on rosenbrock, unstable in despite of initial values (may diverge or find local maxima instead of minima), con: calculation of inverse of hessian - very expensive in large networks
    - Diagonal Hessian (approximation of Newton): $d_i^k \approx \left(\frac{\partial^2 f(x^k)}{(\partial x_i)^2}\right)^{-1}$, very bad performance on Rosenbrock,
    - Gauss Newton method: Too complicated to remember, replace D^k with non linear least square problem, even better performance on rosenbrock then newton, con: again calculation of inverse, but not of hessian
    - **Step size** α:
        - **Minimization rule:** choose α such that $f(x + \alpha d)$ is minimized along d. Hard if f is complicated
        - **Limited minimization rule:** iterative: start small and increase size of α until $f(x)$ is bigger then before, then choose the previous. Easy to implement
        - **Armijo rule:** it is not sufficient that $f(x^{k+1}) < f(x^k)$, thus, the step sizes $\beta^m s$ for $m = 0, 1, ...$ are chosen such that the energy decrease is sufficiently large (dependent on derivation of $f(x)$, formula too complicated), or graphical:

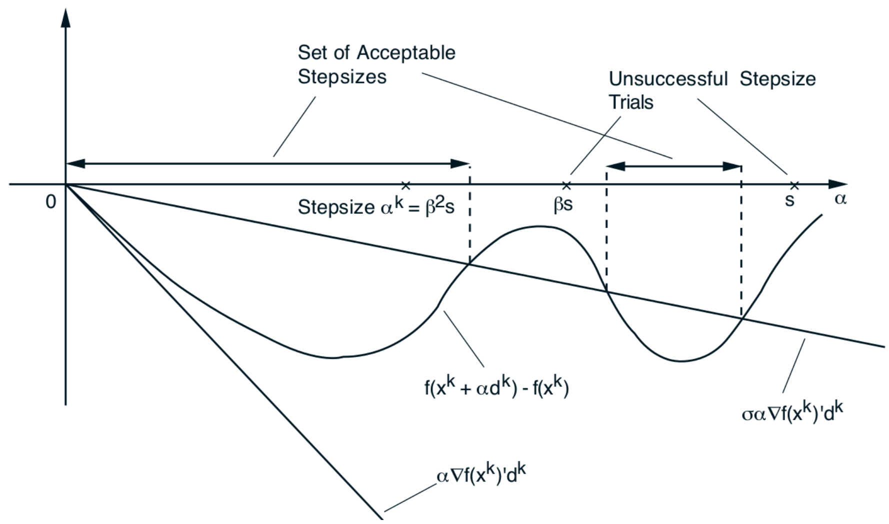

Figure 4: Graphical representation of the idea of Armijo

7. • **Linear:** $\limsup_{k\to\infty} \frac{e(x^{k+1})}{e(x^k)} \leq \beta$ (blue line)

   • **superlinear:** $\limsup_{k\to\infty} \frac{e(x^{k+1})}{e(x^k)^p} < \infty$ (red line)

   • **sublinear:** $\limsup_{k\to\infty} \frac{e(x^{k+1})}{e(x^k)} = 1$ (black line)

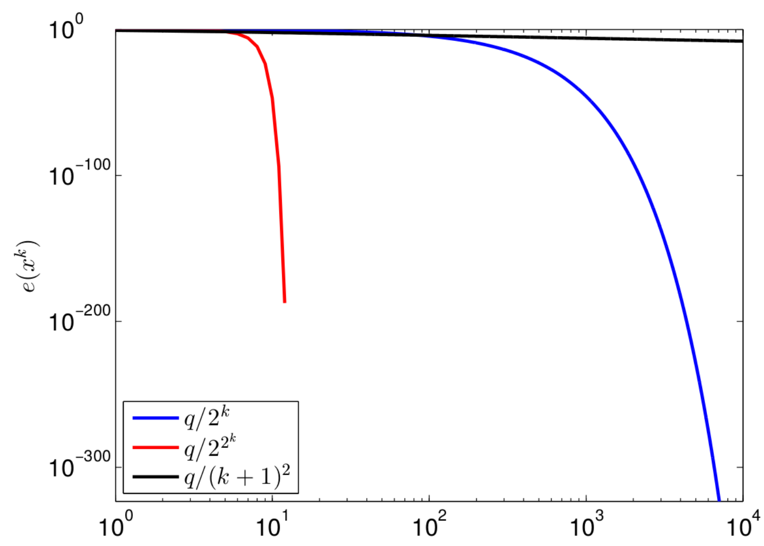

Figure 5: Graphical representation of linear, superinear and sublinear convergence

8. Energy convergence
9. ~~too fast~~
10. ~~polynomial euqations, distance to std. Newton~~
11. incremental of gauss newton
12. ~~too fast~~
13. ~~iterative~~
14. nesterov in gradient, heavy-ball just in point
15. in subspace reduce to eq, what is a subspace?
16. First pages of slide 10
17. middle/end of pages slide 10 - start in interior and just take small steps $->$ we can ignore constraint under these conditions
18. ~~too fast~~
19. see figure 6
20. see figure 7
21. see figure 7

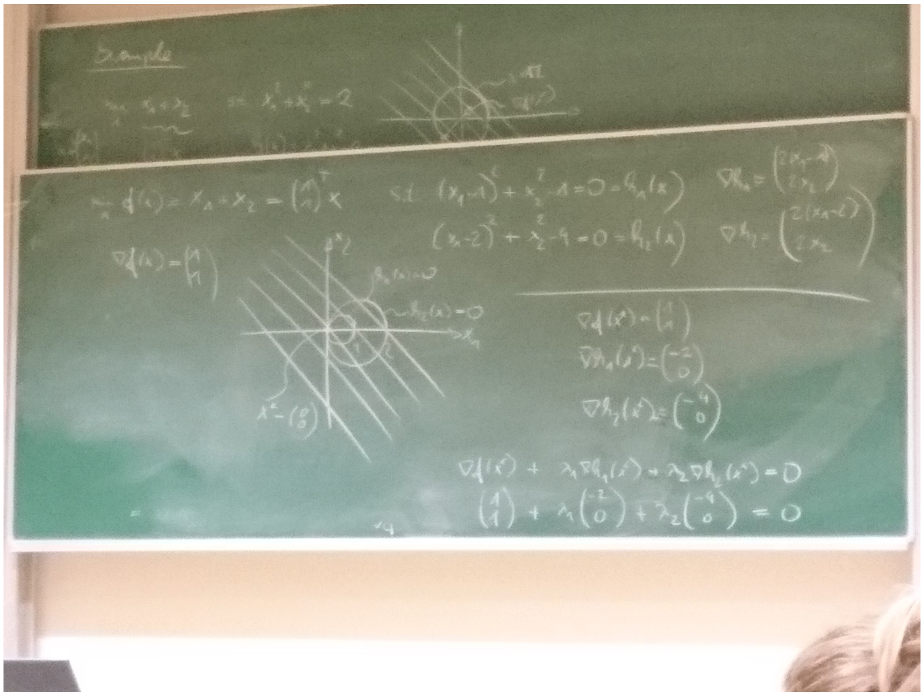

Figure 6: Example1, 24.01.2017

Figure 7: Example 2, 24.01.2017