ECE 5314: Power System Operation & Control

Lecture 12: Gradient and Newton Methods

Vassilis Kekatos

R3 S. Boyd and L. Vandenberghe, *Convex Optimization*, Chapter 9.

R2 A. Gomez-Exposito, A. J. Conejo, C. Canizares, *Electric Energy Systems: Analysis and Operation*, Appendix B.

R1 A. J. Wood, B. F. Wollenberg, and G. B. Sheble, Power Generation, Operation, and Control, Wiley, 2014, Chapter 13.

Unconstrained minimization

Assume f convex, twice continuously differentiable, and finite p^*

$$p^* := \min_{\mathbf{x}} \; f(\mathbf{x})$$

unconstrained minimization methods

- produce sequence of points $\mathbf{x}^t$ with $f(\mathbf{x}^t) \rightarrow p^*$
- interpreted as iterative methods for solving optimality condition

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

- if $\nabla^2 f(\mathbf{x}) \succeq m\mathbf{I}$ with $m > 0$ (strong convexity), then

$$0 \leq f(\mathbf{x}) - p^* \leq \frac{1}{2m}\|\nabla f(\mathbf{x})\|_2^2$$

  useful as stopping criterion (assuming m is known)

Examples

Example 1: unconstrained QP $(\mathbf{P} = \mathbf{P}^\top \succ \mathbf{0})$:

$$\min_{\mathbf{x}} \ \mathbf{x}^\top \mathbf{P} \mathbf{x} + 2\mathbf{q}^\top \mathbf{x} + r$$

Example 2: analytic center of linear inequalities

$$\min_{\mathbf{x}} - \sum_{i=1}^{m} \log(b_i - \mathbf{a}_i^\top \mathbf{x})$$

Example 3: interior-point methods tackle constrained problems by solving a sequence of unconstrained minimization problems

Descent method

1. Compute search direction $\Delta\mathbf{x}^t$
2. Choose step size $\mu_t > 0$
3. Update $\mathbf{x}^{t+1} = \mathbf{x}^t + \mu_t \Delta\mathbf{x}^t$
4. Iterate $(t \rightarrow t+1)$ until stopping criterion is satisfied

Definition: An iterative method is a **descent method** if $f(\mathbf{x}^{t+1}) < f(\mathbf{x}^t)\ \forall t$

Recall for convex f, we have $f(\mathbf{x}^{t+1}) \geq f(\mathbf{x}^t) + (\nabla f(\mathbf{x}))^\top (\mathbf{x}^{t+1} - \mathbf{x}^t)$. Then:

$$f(\mathbf{x}^{t+1}) < f(\mathbf{x}^t) \quad \Rightarrow \quad \text{descent direction satisfies} \quad (\nabla f(\mathbf{x}^t))^\top \Delta\mathbf{x}^t < 0$$

Step size $\mu_t > 0$: constant, exact line search, or backtracking search

$$\text{exact line search :} \quad \mu_t := \arg\min_{\mu > 0} f(\mathbf{x}^t + \mu \Delta\mathbf{x}^t)$$

Gradient descent

1. Compute search direction $\Delta\mathbf{x}^t = -\nabla f(\mathbf{x}^t)$

   (special case of descent method)

2. Choose a step size $\mu_t > 0$

3. Update $\mathbf{x}^{t+1} = \mathbf{x}^t + \mu_t \Delta\mathbf{x}^t$

4. Iterate until stopping criterion is satisfied

- converges with exact or backtracking line search and upper bounded μ
- **convergence rate results**: $c \in (0,1)$ depends on m, $\mathbf{x}^0$, and line search

  linear for strongly convex f: $f(\mathbf{x}^t) - p^* \leq c^t(f(\mathbf{x}^0) - p^*)$

  sublinear for general convex: $f(\mathbf{x}^t) - p^* \leq \frac{L}{t}(f(\mathbf{x}^0) - p^*)$
- very simple but typically slow

Example

$$\min_{\mathbf{x}} \; x_1^2 + M x_2^2$$

where $M > 0$

- exact line search
- initialize at $\mathbf{x}^0 = (M, 1)$

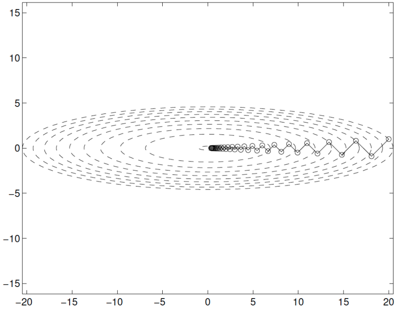


Figure: [Tom Luo's slides]

- iterates take the form

$$\mathbf{x}^t = \left(M \left(\frac{M-1}{M+1} \right)^t, \left(-\frac{M-1}{M+1} \right)^t \right)$$

- fast convergence when M close to 1; one step if $M = 1$!
- slow, zig-zagging if $M \gg 1$ or $M \ll 1$

Example 2

For $m = 100$ and $n = 50$, use gradient method (exact line search)

$$\min_{\mathbf{x}} \; \mathbf{c}^\top \mathbf{x} - \sum_{i=1}^{m} \log(\mathbf{a}_i^\top \mathbf{x} - b_i)$$

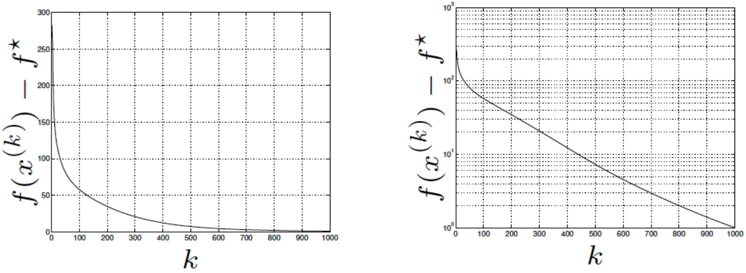


Figure: Function value convergence for gradient method [Z.-Q. Luo's slides]

Steepest descent direction

Term $\nabla f(\mathbf{x})^\top \mathbf{z}$ gives approximate decrease in f for small $\mathbf{z}$

$$f(\mathbf{x}+\mathbf{z}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{z}$$

Find the direction of *steepest descent* (SD):

$$\mathbf{z}_{\text{sd}} = \arg \min_{\|\mathbf{z}\| \leq 1} \nabla f(\mathbf{x})^\top \mathbf{z}$$

Euclidean norm $\|\mathbf{z}\|_2$: $\mathbf{z}_{\text{sd}} = -\nabla f(\mathbf{x}) / \|\nabla f(\mathbf{x})\|_2$ (gradient descent)

Quadratic norm $\|\mathbf{z}\|_{\mathbf{P}} := \sqrt{\mathbf{z}^\top \mathbf{P} \mathbf{z}}$ for some $\mathbf{P} \succeq \mathbf{0}$

$$\mathbf{z}_{\text{sd}} = -\left(\nabla f(\mathbf{x})^\top \mathbf{P}^{-1} \nabla f(\mathbf{x})\right)^{-1/2} \mathbf{P}^{-1} \nabla f(\mathbf{x})$$

Equivalent to SD with Euclidean norm on transformed variables $\mathbf{y} = \mathbf{P}^{1/2}\mathbf{x}$

Geometric interpretation

move as far as possible in direction $-\nabla f(\mathbf{x})$, while staying inside the unit ball

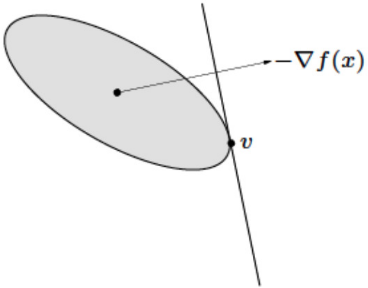


Figure: Boyd's slides

Choosing the norm

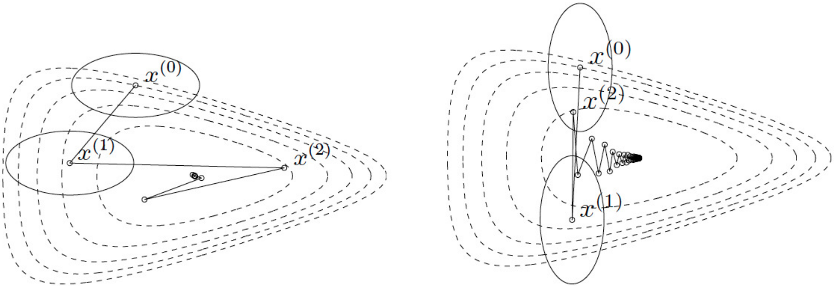


Figure: choice of $\mathbf{P}$ strongly affects speed of convergence [Boyd's slides]

- steepest descent with backtracking line search for two quadratic norms
- ellipses show $\{\mathbf{x} : \ \|\mathbf{x} - \mathbf{x}^t\|_{\mathbf{P}} = 1\}$

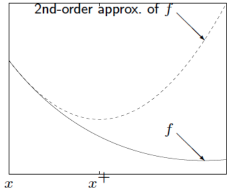
Pure Newton step and interpretations

$$\text{Newton update:} \quad \mathbf{x}^+ = \mathbf{x} + \mathbf{v}$$

$$\text{Newton step:} \quad \mathbf{v} = -\nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x})$$

minimizes second-order expansion of f at $\mathbf{x}$

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{x}^+ - \mathbf{x}) + \frac{1}{2}(\mathbf{x}^+ - \mathbf{x})^\top \nabla^2 f(\mathbf{x})(\mathbf{x}^+ - \mathbf{x})$$



solves linearized optimality condition

$$\nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})(\mathbf{x}^+ - \mathbf{x}) = \mathbf{0}$$

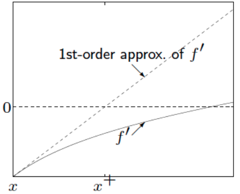
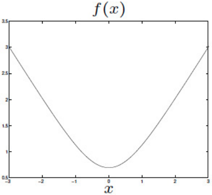
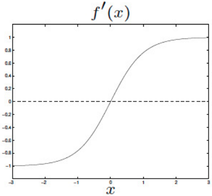


Figure: [Boyd's slides]

Global behavior of Newton iterations

Example: $f(x) = \log(e^x + e^{-x})$, starting at $x^0 = -1.1$





k	$x^{(k)}$	$f(x^{(k)}) - f^\star$
1	$-1.129 \cdot 10^0$	$5.120 \cdot 10^{-1}$
2	$1.234 \cdot 10^0$	$5.349 \cdot 10^{-1}$
3	$-1.695 \cdot 10^0$	$6.223 \cdot 10^{-1}$
4	$5.715 \cdot 10^0$	$1.035 \cdot 10^0$
5	$-2.302 \cdot 10^4$	$2.302 \cdot 10^4$

Figure: pure Newton iterations may diverge! [Z.Q. Luo's slides]

Newton method

Also called *damped or guarded* Newton method

1. Compute Newton direction $\Delta\mathbf{x}^t = -\left[\nabla^2 f(\mathbf{x}^t)\right]^{-1} \nabla f(\mathbf{x}^t)$

2. Choose step size μ_t

3. Update $\mathbf{x}^{t+1} = \mathbf{x}^t + \mu_t \Delta\mathbf{x}^t$

4. Iterate until stopping criterion is satisfied

- **global convergence** with backtracking or exact line search

- quadratic local convergence

- **affine invariance**:

  Newton iterates for $\min_{\mathbf{x}} f(\mathbf{x})$ and $\min_{\mathbf{z}} f(\mathbf{T}\mathbf{z})$ for invertible $\mathbf{T}$

  are equivalent and $\mathbf{x}^t = \mathbf{T}\mathbf{z}^t$

Convergence results

assumptions: $m\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq M\mathbf{I}$ and Lipschitz condition

$$\|\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$$

1. damped Newton phase: $\|\nabla f(\mathbf{x})\|_2 \geq \eta_1$: $f(\mathbf{x}^+) \leq f(\mathbf{x}) - \eta_2$, hence

$$\#\text{iterations} \leq \eta_2^{-1}(f(\mathbf{x}^0) - f^*)$$

2. quadratically convergent phase: $\|\nabla f(\mathbf{x})\|_2 < \eta_1$

$$\#\text{iterations} \leq \log_2 \log_2(\eta_3/\epsilon)$$

total # iterations for reaching accuracy $f(\mathbf{x}^t) - f^* \leq \epsilon$ bounded by:

$$\eta_2^{-1}(f(\mathbf{x}^0) - f^*) + \log_2 \log_2(\eta_3/\epsilon)$$

η_1, η_2, η_3 depend on m, M, L (waived for self-concordant functions)

Example

$$f(\mathbf{x}) = -\sum_{n=1}^{10,000} \log(1 - x_n^2) - \sum_{i=1}^{100,000} \log(b_i - \mathbf{a}_i^\top \mathbf{x})$$

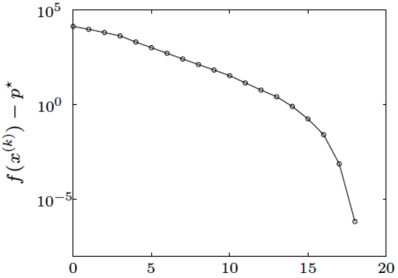


Figure: Two-phase convergence of Newton method [Boyd's slides]

- $\mathbf{x} \in \mathbb{R}^{10,000}$ with sparse $\mathbf{a}_i$'s

Minimization with linear equality constraints

Linearly-constrained optimization problem:

$$\min_{\mathbf{x}} \{f(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

Approach 1: solve *reduced* or *eliminated* problem

$$\min_{\mathbf{z}} f(\mathbf{F}\mathbf{z} + \mathbf{x}_0)$$

where $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$ and $\mathrm{range}(\mathbf{F}) = \mathrm{null}(\mathbf{A})$

Approach 2: Find feasible update that minimizes second-order approximation

$$\Delta\mathbf{x} := \arg\min_{\mathbf{v}} \quad f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{v} + \tfrac{1}{2}\mathbf{v}^\top \nabla^2 f(\mathbf{x})\mathbf{v}$$

$$\text{s.to} \quad \mathbf{A}(\mathbf{x} + \mathbf{v}) = \mathbf{b}$$

[Q: How can this be solved?]