MVE165/MMG630, Applied Optimization
Lecture 12
Unconstrained non-linear programming algorithms and the KKT conditions for constrained nonlinear programs

Ann-Brith Strömberg

2010–04–23

Direction finding methods: Steepest descent (Ch. 10.1)

- Let the search direction be minus the gradient (minimization):

$$\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$$

PROS:

- Requires only gradient information
- Not so computationally demanding per iteration

CONS:

- (Very) Slow convergence towards a stationary point
- Each direction $\mathbf{d}^k$ is perpendicular to the previous one $\mathbf{d}^{k-1}$ (if the line search (i.e., the steplength optimization problem) is solved exactly)—the iterate sequence is "zig-zagging"

Direction finding methods: Newton's method (Ch. 10.2)

- Make use of second derivative information (curvature).
- Requires f to be twice continuously differentiable.
- Taylor expansion of f around $\mathbf{x}$:
  $\varphi_{\mathbf{x}}(\mathbf{d}) := f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{d} + \frac{1}{2}\mathbf{d}^{\mathrm{T}}\nabla^2 f(\mathbf{x})\mathbf{d}$ $(\approx f(\mathbf{x}+\mathbf{d}))$
- We wish to find a direction $\mathbf{d} \in \Re^n$ such that
  $\nabla_{\mathbf{d}}\varphi_{\mathbf{x}}(\mathbf{d}) = \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})\mathbf{d} = \nabla f(\mathbf{x}) + H_f(\mathbf{x})\mathbf{d} = \mathbf{0}^n$
  (a stationary point for $\varphi_{\mathbf{x}}$) $\Rightarrow$ $\mathbf{d}^k = -\mathbf{H}_f(\mathbf{x}^k)^{-1}\nabla f(\mathbf{x}^k)$
- Observe that line search not needed, $t = 1$ (unit step)
- Only look for stationary points for $\varphi_{\mathbf{x}}$ $\Rightarrow$ same $\mathbf{d}^k$ for min/max problems
- If f is quadratic (i.e., $f(\mathbf{x}) = a + \mathbf{c}^{\mathrm{T}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x}$), then Newtons method finds a stationary point for f in one iteration. Verify this!

Direction finding methods: Newton's method

PROS:

- Fast convergence

CONS:

- Convergens towards a stationary point only guaranteed if starting "sufficiently close" to one
- If f is convex around the starting point $\mathbf{x}$ (i.e., $H_f(\mathbf{x})$ positive definite), then Newton's method converges towards a local minimum
- Newton does *not* distinguish between different types of stationary points
- Requires more computations per iteration (matrix inversions)
- Does not always work (if $\mathtt{det}(\mathbf{H}_f(\mathbf{x}^k)) = 0$)

Practical modifications of Newton's method (minimization) (Ch. 10.3)

- Start using Steepest descent, then change to Newton's method
- Use $\mathbf{d}^k = -\mathbf{M}(\mathbf{x}^k)\nabla f(\mathbf{x}^k)$, where $\mathbf{M}(\mathbf{x}^k) \approx \mathbf{H}_f(\mathbf{x}^k)^{-1}$ and $\mathbf{M}(\mathbf{x}^k)$ is positive definite (Quasi-Newton)
- Efficient updates of the inverse should be used
- Let $\mathbf{M}(\mathbf{x}^k) = \left(\mathbf{H}_f(\mathbf{x}^k) + \mathbf{E}^k\right)^{-1}$ such that $\mathbf{M}(\mathbf{x}^k)$ becomes positive definite, e.g., $\mathbf{E}^k = \gamma^k \mathbf{I}$ (which shifts all the eigenvalues by $+\gamma^k$)
- This is called the *Levenberg-Marquardt modification*
- Note: for large values of γ^k, this makes $\mathbf{d}^k$ resemble the steepest descent direction

Optimization over convex sets

Up to now, we have looked at unconstrained optimization. Now:

$$\text{minimize } f(\mathbf{x}) \text{ subject to } \mathbf{x} \in S$$

where $S = \{ \mathbf{x} \in \Re^n \mid g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m \}$ is a convex set

- **Definition** Feasible direction
  If $\mathbf{x} \in S$, then $\mathbf{d} \in \Re^n$ is a feasible direction from $\mathbf{x}$ if a small step in this direction does not lead outside the set S (cf. the simplex method for LP)

  Formally: $\mathbf{d}$ defines a feasible direction at $\mathbf{x} \in S$ if

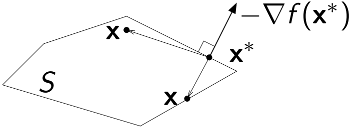
  $$\exists \delta > 0 \text{ such that } \mathbf{x} + t\mathbf{d} \in S \text{ for all } t \in [0, \delta]$$

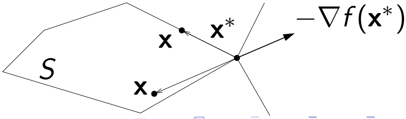
- **Definition** Active constraints
  The active constraints at $\mathbf{x} \in S$ are those that are fulfilled with equality, i.e., $\mathcal{I}(\mathbf{x}) = \{ i = 1, \ldots, m \mid g_i(\mathbf{x}) = 0 \}$
- Draw!!

Optimality conditions (Ch. 11)

- **Definition** FEASIBLE DIRECTIONS FOR LINEAR CONSTRAINTS
  Suppose that $g_i(\mathbf{x}) = \mathbf{a}_i^{\mathrm{T}}\mathbf{x} - b_i$, $i = 1, \ldots, m$. Then, the set of feasible directions at $\mathbf{x}$ is $\{\, \mathbf{d} \in \Re^n \mid \mathbf{a}_i^{\mathrm{T}}\mathbf{d} \leq 0, i \in \mathcal{I}(\mathbf{x}) \,\}$
- **Necessary optimality conditions**
  If $\mathbf{x}^* \in S$ is a local minimum of f over S then $\nabla f(\mathbf{x}^*)^{\mathrm{T}}\mathbf{d} \geq 0$ holds for all feasible directions $\mathbf{d}$ at $\mathbf{x}^*$
  (i.e., at $\mathbf{x}^*$ there are no feasible descent directions)
- **Necessary and sufficient optimality conditions**
  Suppose S is non-empty and convex and f convex. Then,

$$\begin{aligned} &\mathbf{x}^* \text{ is a global minimum of } f \text{ over } S \\ \Leftrightarrow\ &\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) \geq 0 \text{ holds for all } \mathbf{x} \in S \end{aligned}$$



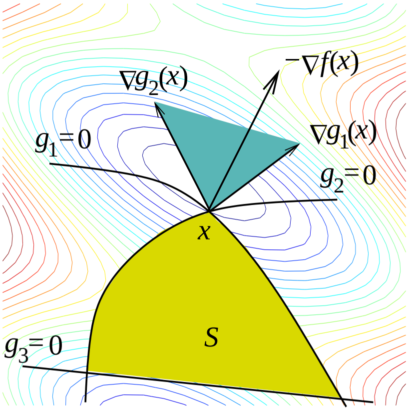


The Karush-Kuhn-Tucker conditions: Necessary conditions for optimality

- Assume that the functions $g_i : \Re^n \mapsto \Re$, $i = 1, \ldots, m$, are convex and differentiable and that there exists a point $\bar{\mathbf{x}} \in S$ such that $g_i(\bar{\mathbf{x}}) < 0$, $i = 1, \ldots, m$.
- Further, assume that $f : \Re^n \mapsto \Re$ is differentiable.
- If $\mathbf{x}^* \in S$ is a local minimum of f over S, then there exists a vector $\boldsymbol{\mu} \in \Re^m$ such that

$$
\begin{aligned}
\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\mathbf{x}^*) &= \mathbf{0}^n \\
\mu_i g_i(\mathbf{x}^*) &= 0, \qquad i = 1, \ldots, m \\
g_i(\mathbf{x}^*) &\leq 0, \qquad i = 1, \ldots, m \\
\boldsymbol{\mu} &\geq \mathbf{0}^m
\end{aligned}
$$

Geometry of the Karush-Kuhn-Tucker conditions



Figur: Geometric interpretation of the Karush-Kuhn-Tucker conditions. At a local minimum, minus the gradient of the objective can be expressed as a non-negative linear combination of the gradients of the active constraints at this point.

The Karush-Kuhn-Tucker conditions: Sufficient conditions under convexity

- Assume that the functions $f, g_i : \Re^n \mapsto \Re$, $i = 1, \ldots, m$, are convex and differentiable.
- If the conditions

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\mathbf{x}^*) &= \mathbf{0}^n \\ \mu_i g_i(\mathbf{x}^*) &= 0, \qquad i = 1, \ldots, m \\ \boldsymbol{\mu} &\geq \mathbf{0}^m \end{aligned}$$

  hold, then $\mathbf{x}^* \in S$ is a global minimum of f over $S = \{\, \mathbf{x} \in \Re^n \mid g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m \,\}$.
- The Karush-Kuhn-Tucker conditions can also be stated for optimization problems with equality constraints

The optimality conditions can be used to..

- verify an (local) optimal solution
- solve certain special cases of nonlinear programs (e.g. quadratic)
- algorithm construction
- derive properties of a solution to a non-linear program

Example

$$\begin{array}{lrcl} \text{minimize} & f(\mathbf{x}) := 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} & x_1^2 + x_2^2 & \leq & 5 \\ & 3x_1 + x_2 & \leq & 6 \end{array}$$

- Is $\mathbf{x}^0 = (1, 2)^{\mathrm{T}}$ a Karush-Kuhn-Tucker point?
- An optimal solution?
- $\nabla f(\mathbf{x}) = (4x_1 + 2x_2 - 10, 2x_1 + 2x_2 - 10)^{\mathrm{T}}$,
  $\nabla g_1(\mathbf{x}) = (2x_1, 2x_2)^{\mathrm{T}}$, $\nabla g_2(\mathbf{x}) = (3, 1)^{\mathrm{T}}$

$$\Rightarrow \begin{bmatrix} 4x_1^0 + 2x_2^0 - 10 + 2x_1^0\mu_1 + 3\mu_2 = 0 \\ 2x_1^0 + 2x_2^0 - 10 + 2x_2^0\mu_1 + \mu_2 = 0 \\ \mu_1((x_1^0)^2 + (x_2^0)^2 - 5) = \mu_2(3x_1^0 + x_2^0 - 6) = 0 \\ \mu_1, \mu_2 \geq 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 2\mu_1 + 3\mu_2 = 2 \\ 4\mu_1 + \mu_2 = 4 \\ 0\mu_1 = -\mu_2 = 0 \\ \mu_1, \mu_2 \geq 0 \end{bmatrix}$$

$$\Rightarrow \quad \mu_2 = 0 \quad \Rightarrow \quad \mu_1 = 1 \geq 0$$

Example, continued

- The Karush-Kuhn-Tucker conditions hold.
- Optimal? Check convexity!
- $\nabla^2 f(\mathbf{x}) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$, $\nabla^2 g_1(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\nabla^2 g_2(\mathbf{x}) = \mathbf{0}^{2\times 2}$

$\Rightarrow$ f, g_1, and g_2 are convex $\Rightarrow$ $\mathbf{x}^0 = (1,2)^{\mathrm{T}}$ is an optimal solution $f(\mathbf{x}^0) = -20$