Constrained optimization: direct methods (cont.)

Jussi Hakanen
Post-doctoral researcher
jussi.hakanen@jyu.fi

spring 2014 TIES483 Nonlinear optimization
Direct methods

Also known as *methods of feasible directions*

Idea

– in a point $x^h$, generate a feasible search direction where objection function value can be improved
– use line search to get $x^{h+1}$

Methods differ in

– how to choose a feasible direction and
– what is assumed from the constraints (linear/nonlinear, equality/inequality)
Examples of direct methods

- Projected gradient method
- Active set method
- Sequential Quadratic Programming (SQP) method
Active set method

Consider a problem \( \min f(x) \) s.t. \( Ax \leq b \), where \( A \) is an \( m \times n \) matrix and \( b \in \mathbb{R}^m \)
- \( x^* \in S \) if and only if \( Ax^* \leq b \)
- \( i \)th constraint: \( (a^i)^T x \leq b_i \)

Idea
- In \( x^h \), the set of constraints is divided into active (\( i \in I \)) and inactive constraints
- Inactive constraints are not taken into account when the search direction \( d^h \) is determined
- Inactive constraints affect only when computing the optimal step length \( \alpha^h \)
Feasible directions

- For $i \in I$, $(a^i)^T x^h = b_i$
- If $d^h$ is feasible in $x^h$, then $x^h + \alpha d^h \in S$ for some $\alpha > 0$ → 
  $$(a^i)^T (x^h + \alpha d^h) = (a^i)^T x^h + \alpha (a^i)^T d^h \leq b_i$$
- $(a^i)^T d^h \leq 0$ for feasible $d^h$ and the constraint remains active if $(a^i)^T d^h = 0$
Optimality conditions

Assume that \( t \) constraints are active in \( x^h \) (\( I = \{i_1, \ldots, i_t\} \)) and denote by \( \bar{A} \) the \( t \times n \) matrix that consists of the active constraints and \( \bar{Z} \) is the matrix spanning \( \{d \mid \bar{A}d = 0\} \)

**Necessary:** If \( x^* \in S \) \( (\bar{A}x^* = \bar{b} = (b_{i_1}, \ldots, b_{i_t})) \) is a local minimizer, then

1) \( \bar{Z}^T \nabla f(x^*) = 0 \) that is \( \nabla f(x^*) + \bar{A}^T \bar{\mu}^* = 0 \),
2) \( \mu_i^* = \bar{\mu}_i^* > 0 \) for all \( i \in I \) and
3) matrix \( \bar{Z}^T H(X^*)\bar{Z} \) is positive semidefinite

**Sufficient:** If in \( x^* \in S \)

1) \( \bar{Z}^T \nabla f(x^*) = 0 \),
2) \( \mu_i^* = \bar{\mu}_i^* > 0 \) for all \( i \in I \) and
3) matrix \( \bar{Z}^T H(x^*)\bar{Z} \) is positive definite,
then \( x^* \) is a local minimizer
On active constraints

- Optimization problem with inequality constraints is more difficult than problem with equality constraints since the active set in a local minimizer is not known.

- If it would be known, then it would be enough to solve a corresponding equality constrained problem.

- In that case, if the other constraints would be satisfied in the solution and all the Lagrange multipliers were non-negative, then the solution would also be a solution to the original problem.
Using the active set

- At each iteration, a working set is considered which consists of the active constraints in $x^h$
- The direction $d^h$ is determined so that it is a descent direction in the working set
  - E.g. Rosen’s projected gradient method can be used
Algorithm

1) Choose a starting point $x^1$. Determine the active set $I^1$ and the corresponding matrices $A^1$ and $Z^1$ in $x^1$. Set $h = 1$.

2) Compute a feasible direction $d^h$ in the subspace defined by the active constraints. (E.g. by using projected gradient.)

3) If $\|d^h\| = 0$, go to 6). Otherwise, find $\alpha_{max} = \max\{\alpha \mid x^h + \alpha d^h \in S\}$ and solve
   \[
   \min f(x^h + \alpha d^h) \text{ s.t. } 0 \leq \alpha \leq \alpha_{max}.
   \]
   Let the solution be $\alpha_h$. Set $x^{h+1} = x^h + \alpha_h d^h$.

4) If $\alpha_h < \alpha_{max}$, go to 7). (Active set has changed.)

5) Addition to active set: Add the constraint which first becomes active to $I^h$. Update $A^h$ and $Z^h$ correspondingly $\rightarrow I^{h+1}$, $A^{h+1}$ and $Z^{h+1}$. Go to 7).

6) Removal from active set: Compute $\mu_i, i \in I^h$. If $\mu_i \geq 0$ for all $i \in I^h$, $x^h$ is optimal, stop. If $\mu_i < 0$ for some $i$, remove the corresponding constraint from active set (the one with the smallest $\mu_i$). Update $A^h$ and $Z^h$ correspondingly $\rightarrow I^{h+1}$, $A^{h+1}$ and $Z^{h+1}$. Set $x^{h+1} = x^h$.

7) Set $h = h + 1$ and go to 2).
How to compute $\mu$

- At each iteration, a problem
  \[
  \min f(x) \text{ s.t. } Ax = b
  \]
  is solved
- Lagrangian function: $L(x, \mu) = f(x) + \mu^T(\bar{A}x - \bar{b})$
- In a minimizer $x^*$, $\nabla L(x^*, \mu) = \nabla f(x^*) + \bar{A}^T \mu = 0$
  \[
  \Rightarrow \nabla f(x^*) = -\bar{A}^T \mu
  \]
  \[
  \Rightarrow \bar{A} \nabla f(x^*) = -(\bar{A} \bar{A}^T) \mu
  \]
  \[
  \Rightarrow \mu = -(\bar{A} \bar{A}^T)^{-1} \bar{A} \nabla f(x^*)
  \]
- Used in $x^h$ to approximate $\mu$
Example: Active set method + Rosen’s projected gradient

Minimise \[ f(x) = (x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 - 3)^2 + (x_4 - 4)^2 \]

Subject to:
1. \[ x_1 + x_2 + x_3 + x_4 \leq 5 \]
2. \[ 3x_1 + 3x_2 + 2x_3 + x_4 \leq 10 \]
3. \[ -x_1 \leq 0 \]
4. \[ -x_2 \leq 0 \]
5. \[ -x_3 \leq 0 \]
6. \[ -x_4 \leq 0 \]

1. Select an initial point \( x^1 = (\frac{1}{2}, 1, \frac{3}{2}, 2)^T \). The active set is \( I^1 = \{1\} \) and \( \bar{A}^1 = (1 \ 1 \ 1 \ 1) \).
2. Compute the search direction. We use the projected gradient, which in this case is given by \( P = I - (\bar{A}^1)^T(\bar{A}^1(\bar{A}^1)^T)^{-1}\bar{A}^1 \). Thus, \( (\bar{A}^1(\bar{A}^1)^T)^{-1} = \frac{1}{4} \) and

\[
P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}.
\]
Example (cont.)

Nyt $\nabla f(x^1) = (-1, -2, -3, -4)^T$, joten hakusuunta on

$$d^1 = -P \nabla f(x^1) = \left( -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2} \right)^T.$$  

3. Etsitään suurin $\lambda$, jolla $x^1 + \lambda d^1$ on sallittu. Rajoitteista saadaan ehdot $\lambda \geq -0.143$, $\lambda \leq \frac{1}{3}$, $\lambda \leq 2$, $\lambda \geq -3$ ja $\lambda \geq -\frac{4}{3}$. Suurin $\lambda$, joille nämä kaikki toteutuvat on $\lambda_{max} = \frac{1}{3}$. Tehtävän

minimoi $f(x^1 + \lambda d^1)$
rajoittein $0 \leq \lambda \leq \lambda_{max}$

ratkaisu on $\lambda_1 = \frac{1}{3}$ (rajoittamaton optimi on $\lambda = \frac{1}{2}$). Nyt on $f(x^1 + \lambda_1 d^1) < f(x^1)$.

Asetetaan $x^2 = x^1 + \lambda_1 d^1 = (0, \frac{5}{6}, \frac{5}{3}, \frac{5}{2})^T$.

5. Aktiivisten rajoitteiden joukkoon tuli yksi rajoite lisää, joten $I^2 = \{1, 3\}$. Vastaavasti

$$\tilde{A}^2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$  

7. Lisätään iteraatiolaskuria $h = 2$. - - - - - - -
Example (cont.)

2. Lasketaan uusi projektiomatriisi \( P = I - (A^2)(A^2)^{-1}A^2 \) eli
\[
P = I - \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 4/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 2 \\ 0 & -1 & -1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix}.
\]
Gradientti on \( \nabla f(x^2) = (-2, -\frac{7}{3}, -\frac{8}{3}, -3)^T \). Uusi hakusuunta on
\[
d^2 = -P\nabla f(x^2) = \left(0, -\frac{1}{3}, 0, \frac{1}{3}\right)^T.
\]

3. Etsitään suurin \( \lambda \), jolla \( x^2 + \lambda d^2 \) on sallittu. Rajoitteista saadaan ehdot \( \lambda \geq -2.5 \), \( \lambda \leq 2.5 \) ja \( \lambda \geq -7.5 \). Suurin \( \lambda \), joille nämä kaikki toteutuvat on \( \lambda_{\text{max}} = 2.5 \). Tehtävän

minimoi \( f(x^2 + \lambda d^2) \)
rajoittein \( 0 \leq \lambda \leq \lambda_{\text{max}} \)
ratkaisu on \( \lambda_2 = \frac{1}{2} \). Nyt on \( f(x^2 + \lambda_2 d^2) < f(x^2) \). Aktiivisten rajoitteiden joukko ei muuttunut. Saadaan \( x^3 = x^2 + \lambda_2 d^2 = (0, \frac{2}{3}, \frac{5}{3}, \frac{8}{3})^T \).

7. Lisätään laskuria \( h = 3 \).
2. Projektiomatriisi \( P \) on siis entinen ja \( \nabla f(x^3) = (-2, -\frac{8}{3}, -\frac{8}{3}, -\frac{8}{3})^T \). Uusi hakusuunta on \( d^3 = -P\nabla f(x^3) = (0, 0, 0, 0)^T \).
3. Nyt on \( \|d^3\| = 0 \).
6. Lasketaan Lagrangen kertoimet \( \tilde{\mu} = -(A^2)^{-1}A^2\nabla f(x^3) = \left(\frac{8}{3}, \frac{2}{3}\right)^T \). Nyt on siis \( \tilde{\mu}_i \geq 0 \) kaikilla \( i \in I^2 \), joten minimipiste on \( x^3 \) ja voidaan lopettaa.
SQP method

- **Sequential Quadratic Programming**
- Idea is to generate a sequence of quadratic optimization problems whose solutions approach the solution of the original problem
- Quadratic problems are based on applying KKT conditions to the original problem
  - Minimize a quadratic approximation of the Lagrangian function with respect to linear approximation of the constraints
  - Also referred as *projected Lagrangian method*
Approximation of the constraints

Consider a problem

\[ \min f(x) \text{ s.t. } h_i(x) = 0, i = 1, \ldots, l, \]

where all the functions are twice continuously differentiable.

Taylor’s series \((d = x^* - x^h)\):

\[ h_i(x^h + d) \approx h_i(x^h) + \nabla h_i(x^h)^T d \]

\[ h_i(x^*) = 0 \text{ for all } i \]

\[ \implies \nabla h(x^h)d = -h(x^h) \]
Approximation of the Lagrangian

\[ L(x^h + d, \nu^*) \approx L(x^h, \nu^*) + d^T \nabla_x L(x^h, \nu^*) + \frac{1}{2} d^T \nabla^2_{xx} L(x^h, \nu^*) d \]

A quadratic subproblem:

\[
\begin{align*}
\min_d & \quad d^T \nabla f(x^h) + \frac{1}{2} d^T E(x^h, \nu^h) d \\
\text{s. t.} & \quad \nabla h(x^h) d = -h(x^h),
\end{align*}
\]

where \( E(x^h, \nu^h) \) is either the Hessian of the Lagrangian or some approximation of it

– Approximate if the second derivatives are not available

It can be shown (under some assumptions), that solutions of the subproblems approach \( x^* \) and Lagrange multipliers approach \( \nu^* \)
Algorithm

1) Choose a starting point $x^1$. Compute $E^1$ and set $h = 1$.

2) Solve a quadratic subproblem
\[
\min_{d} d^T \nabla f(x^h) + \frac{1}{2} d^T E^h d \quad \text{s.t.} \quad \nabla h(x^h) d = -h(x^h).
\]
Let the solution be $d^h$.

3) Set $x^{h+1} = x^h + d^h$. Stop if $x^{h+1}$ satisfies optimality conditions.

4) Compute an estimate for $\nu^{h+1}$ and compute $E^{h+1} = E(x^{h+1}, \nu^{h+1})$.

5) Set $h = h + 1$ and go to 2).
Notes

- Quadratic subproblems are typically not solved as optimization problems but are converted into a system of equations and solved with suitable methods:
  \[ E^h d = 0, \]
  \[ \nabla h(x^*) d = 0 \]

- If \( E = \nabla^2_{xx} L \), then \( (x^h, v^h) \rightarrow (x^*, v^*) \) quadratically

- If \( E \) is an approximation, then convergence is superlinear

- Only local convergence, that is, starting point must be 'close' to optimum

- Inequality constraints considered with active set strategies or explicitly as \( \nabla g_i(x^h)^T d \leq -g_i(x^h) \)