Lecture 12, nonlinear programming

Ann-Brith Strömberg

29 April 2008

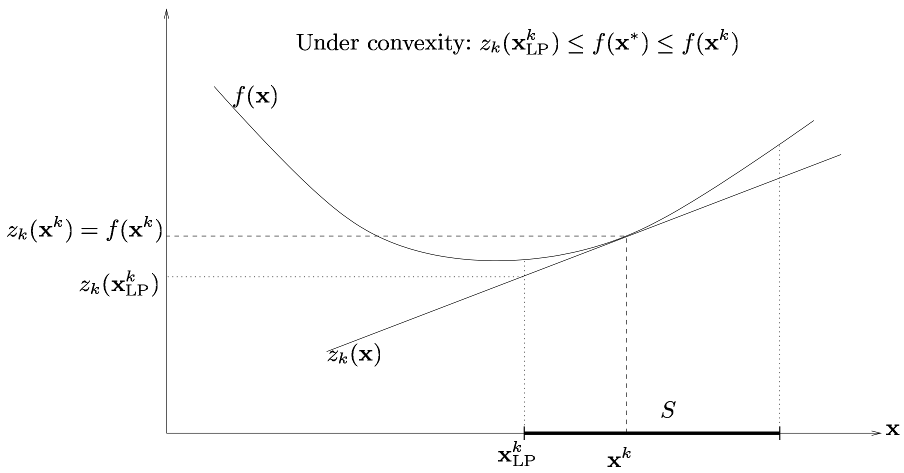
The Frank-Wolfe algorithm for minimizing a nonlinear function over a polyhedral feasible set

Assume: f convex, S bounded polyhedron: $\begin{bmatrix} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in S \end{bmatrix}$

1. Choose $\mathbf{x}^0 \in S$ and $\varepsilon > 0$. Let $\mathtt{UB} = f(\mathbf{x}^0)$, $\mathtt{LB} = -\infty$, $k = 0$.
2. Solve the linear approximation (LP):

   $$\min_{\mathbf{x} \in S} z_k(\mathbf{x}) := f(\mathbf{x}^k) + \nabla f(\mathbf{x}^k)^T(\mathbf{x} - \mathbf{x}^k) \quad \Rightarrow \quad \mathbf{x} = \mathbf{x}^k_{\mathrm{LP}}$$

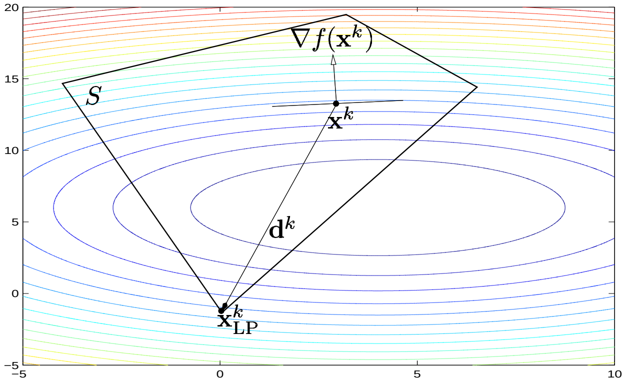
   Let $\mathbf{d}^k = \mathbf{x}^k_{\mathrm{LP}} - \mathbf{x}^k$, $\mathtt{LB} = \max\{\mathtt{LB}, z_k(\mathbf{x}^k_{\mathrm{LP}})\}$. Stop if $\mathtt{UB} - \mathtt{LB} < \varepsilon$.
3. Solve $\min_{0 \le t \le 1} \varphi(t) := f(\mathbf{x}^k + t \cdot \mathbf{d}^k) \quad \Rightarrow \quad t = t_k$
4. Let $\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \mathbf{d}^k$, $\mathtt{UB} = f(\mathbf{x}^{k+1})$
5. Stop if $\mathtt{UB} - \mathtt{LB} < \varepsilon$. Otherwise, $k := k + 1$, go to step 2



Figur 1: Illustration of the Frank–Wolfe algorithm

The Frank–Wolfe-algorithm

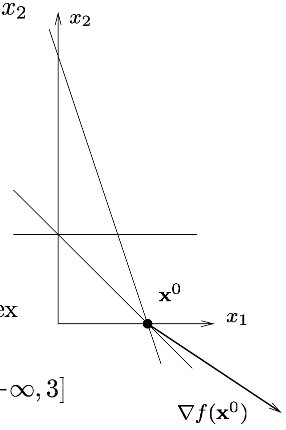
- Solves a non-linear optimization problem using a *sequence of approximating*, linear (easier) *problems*, and a sequence of one dimensional (easy) non-linear problems.
- *Estimates* of the optimal objective value is used to terminate the procedure at a *guaranteed maximal deviation* from an optimal solution ($\varepsilon > 0$).



Figur 2: Step 1 of the Frank–Wolfe algorithm.

An example solved by the the Frank–Wolfe-algorithm

$$\begin{array}{llrcl} \text{minimize} & f(\mathbf{x}) = & 3x_1^2 + x_2^2 - x_1x_2 - 3x_2 \\ \text{subject to} & & x_1 + x_2 & \geq & 1 \\ & & 3x_1 + x_2 & \leq & 3 \\ & & x_2 & \leq & 1 \end{array}$$



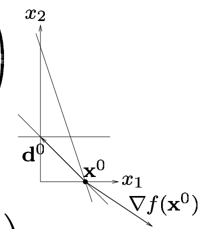
$$\nabla f(\mathbf{x}) = \begin{pmatrix} 6x_1 - x_2 \\ 2x_2 - x_1 - 3 \end{pmatrix}$$

$$\mathbf{H}_f(\mathbf{x}) = \begin{pmatrix} 6 & -1 \\ -1 & 2 \end{pmatrix} \text{ positive definite } \Rightarrow f \text{ strictly convex}$$

$$\mathbf{x}^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; f(\mathbf{x}^0) = 3 \quad \Rightarrow \quad [\mathrm{LB}, \mathrm{UB}] = [-\infty, 3]$$

$$z_0(\mathbf{x}) = 6x_1 - 4x_2 - 3 \Rightarrow \mathbf{x}^0_{\mathrm{LP}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; \mathbf{d}^0 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$z_0(\mathbf{x}^0_{\mathrm{LP}}) = -7 \quad \Rightarrow \quad [\mathrm{LB}, \mathrm{UB}] = [-7, 3]$$

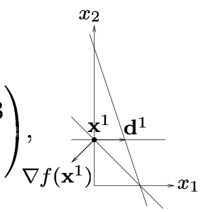


$$\left.\begin{array}{l} \mathbf{x}^0 + t \cdot \mathbf{d}^0 = \begin{pmatrix} 1-t \\ t \end{pmatrix} \\ \varphi(t) = 3(1-t)^2 + t^2 - (1-t)t - 3t \\ \varphi'(t) = 10t - 10 = 0 \Rightarrow t_0 = 1 \end{array}\right\} \Rightarrow \mathbf{x}^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

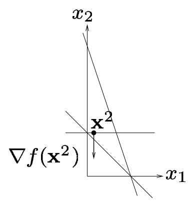
$$f(\mathbf{x}^1) = -2 \quad \Rightarrow \quad [\mathrm{LB}, \mathrm{UB}] = [-7, -2]$$

$$z_1(\mathbf{x}) = -x_1 - x_2 - 1 \Rightarrow \mathbf{x}^1_{\mathrm{LP}} = \begin{pmatrix} 2/3 \\ 1 \end{pmatrix}, \mathbf{d}^1 = \begin{pmatrix} 2/3 \\ 0 \end{pmatrix},$$

$$z_1(\mathbf{x}^1_{\mathrm{LP}}) = -8/3 \quad \Rightarrow \quad [\mathrm{LB}, \mathrm{UB}] = [-8/3, -2]$$



$$\left.\begin{array}{l} \mathbf{x}^1 + t \cdot \mathbf{d}^1 = \begin{pmatrix} 2t/3 \\ 1 \end{pmatrix} \\ \varphi(t) = 4t^2/3 - 2t/3 - 2 \\ \varphi'(t) = 8t/3 - 2/3 = 0 \\ \Rightarrow \quad t_1 = 1/4 \end{array}\right\} \Rightarrow \mathbf{x}^2 = \begin{pmatrix} 1/6 \\ 1 \end{pmatrix}$$



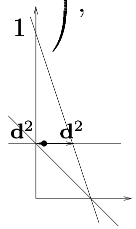
$$f(\mathbf{x}^2) = -25/12 \quad \Rightarrow \quad [\mathrm{LB}, \mathrm{UB}] = [-8/3, -25/12]$$

$$z_2(\mathbf{x}) = -7x_2/6 - 11/12 \quad \Rightarrow \quad \mathbf{x}^2_{\mathrm{LP}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ alt. } \begin{pmatrix} 2/3 \\ 1 \end{pmatrix},$$

$$z_2(\mathbf{x}^2_{\mathrm{LP}}) = -25/12 \quad \Rightarrow \quad [\mathrm{LB}, \mathrm{UB}] = [-25/12, -25/12]$$

$\Rightarrow$ Optimum!

$$\mathbf{x}^* = \mathbf{x}^2 = \begin{pmatrix} 1/6 \\ 1 \end{pmatrix}, \quad f(\mathbf{x}^*) = -25/12$$



Check the Karush-Kuhn-Tucker conditions at $\mathbf{x}^*$

$\mathbf{x}^* = (1/6, 1)^{\mathrm{T}}$

$f(\mathbf{x}) = 3x_1^2 + x_2^2 - x_1x_2 - 3x_2$

$g_1(\mathbf{x}) = -x_1 - x_2 + 1 \leq 0,\ g_2(\mathbf{x}) = 3x_1 + x_2 - 3 \leq 0,\ g_3(\mathbf{x}) = x_2 - 1 \leq 0$

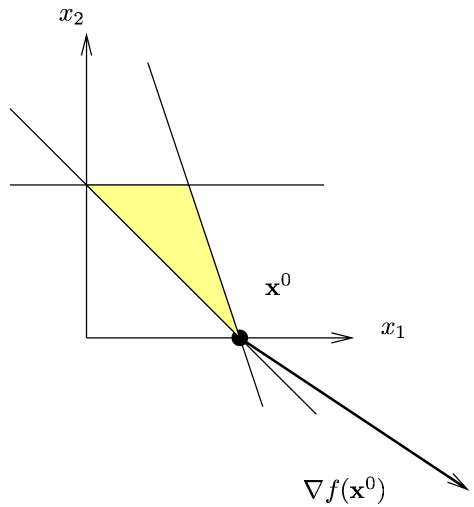
$\nabla f(\mathbf{x}^*) = (0, -7/6)^{\mathrm{T}}$

$\nabla g_1(\mathbf{x}^*) = (-1,-1)^{\mathrm{T}},\ \nabla g_2(\mathbf{x}^*) = (3,1)^{\mathrm{T}},\ \nabla g_3(\mathbf{x}^*) = (0,1)^{\mathrm{T}}$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{3} \mu_i \nabla g_i(\mathbf{x}^*) = (0,0)^{\mathrm{T}} \Rightarrow \begin{cases} 0 - 1\cdot\mu_1 + 3\cdot\mu_2 + 0\cdot\mu_3 = 0 \\ -7/6 - 1\cdot\mu_1 + 1\cdot\mu_2 + 1\cdot\mu_3 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} -\mu_1 + 3\mu_2 = 0 \\ -\mu_1 + \mu_2 + \mu_3 = 7/6 \end{cases}$$

$\mu_i \cdot g_i(\mathbf{x}^*) = 0 \Rightarrow \mu_1 = \mu_2 = 0 \Rightarrow \mu_3 = 7/6 \geq 0 \Rightarrow$ Global optimum



Quadratic programming (QP)

Example (quadratic convex objective, linear constraints):

$$\begin{array}{lrcl} \text{minimize} & f(\mathbf{x}) & = & -2x_1 - 6x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{subject to} & & & x_1 + x_2 \leq 2 \\ & & & -x_1 + 2x_2 \leq 2 \\ & & & x_1,\ x_2 \geq 0 \end{array}$$

Generally:

minimize $\mathbf{q}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} - \mathbf{b} \leq \mathbf{0}, -\mathbf{I}\mathbf{x} \leq \mathbf{0}$

where

$$\mathbf{q} = \begin{pmatrix} -2 \\ -6 \end{pmatrix},\ \mathbf{Q} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix},\ \mathbf{A} = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix},\ \mathbf{b} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},\ \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

QP: The Karush-Kuhn-Tucker conditions

$$\begin{array}{rcl} \mathbf{q} + \mathbf{Q}\mathbf{x} + \mathbf{A}^T\boldsymbol{\mu} - \mathbf{I}\boldsymbol{\lambda} & = & 0 \\ \mathbf{A}\mathbf{x} & \leq & \mathbf{b} \\ -\mathbf{I}\mathbf{x} & \leq & \mathbf{0} \\ \boldsymbol{\mu}, \boldsymbol{\lambda} & \geq & \mathbf{0} \\ \boldsymbol{\mu}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \boldsymbol{\lambda}^T\mathbf{x} & = & 0 \end{array}$$

Slack variables $\mathbf{s} \geq \mathbf{0}$ of the constraints $\mathbf{A}\mathbf{x} \leq \mathbf{b}$: $\quad \mathbf{A}\mathbf{x} + \mathbf{s} = \mathbf{b}$

$\Rightarrow$ The Karush-Kuhn-Tucker constraints reduce to:

$$\begin{array}{rcl} \mathbf{Q}\mathbf{x} + \mathbf{A}^T\boldsymbol{\mu} - \mathbf{I}\boldsymbol{\lambda} & = & -\mathbf{q} \\ \mathbf{A}\mathbf{x} + \mathbf{I}\mathbf{s} & = & \mathbf{b} \\ \mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{s} & \geq & \mathbf{0} \\ \mu_i s_i = \lambda_j x_j & = & 0 \text{ for all } i, j \end{array}$$

QP: The Karush-Kuhn-Tucker conditions

- Convex optimization problem ⇒ Karush-Kuhn-Tucker conditions are sufficient for a global optimum

⇒ A solution $(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{s})$ that fulfils the Karush-Kuhn-Tucker conditions is optimal for the quadratic program (QP)

- The system is linear, with variables: $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \mathbf{s} \geq \mathbf{0}$
- Additional conditions: $\mu_i s_i = \lambda_j x_j = 0$ for all i, j
- Linear programming—Simplex algorithm with *restricted basis*:
- Either $\mu_i = 0$ or $s_i = 0$. Either $\lambda_j = 0$ or $x_j = 0$.

⇒ If, e.g., s_2 is in the basis ($s_2 > 0$), μ_2 may *not* enter the basis

- Introduce artificial variables where needed and solve a Phase 1 problem

The phase 1 problem—example

$$
\begin{array}{llcccccccccccl}
\text{minimize} & w = & & & & & & & & & a_1 & +a_2 & & \\
\text{subject to} & 2x_1 & -2x_2 & +\mu_1 & -\mu_2 & -\lambda_1 & & & & & +a_1 & & = & 2 \\
 & -2x_1 & +4x_2 & +\mu_1 & +2\mu_2 & & -\lambda_2 & & & & & +a_2 & = & 6 \\
 & x_1 & +x_2 & & & & & +s_1 & & & & & = & 2 \\
 & -x_1 & +2x_2 & & & & & & +s_2 & & & & = & 2 \\
 & x_1, & x_2, & \mu_1, & \mu_2, & \lambda_1, & \lambda_2, & s_1, & s_2, & a_1, & a_2 & & \geq & 0 \\
 & \multicolumn{13}{l}{\mu_1 s_1 = 0, \quad \mu_2 s_2 = 0, \quad \lambda_1 x_1 = 0, \quad \lambda_2 x_2 = 0}
\end{array}
$$

Find a starting base by reformulating: $a_1, a_2, s_1, s_2 \Rightarrow$
$w - a_1 - a_2 = w + 2x_2 + 2\lambda_1 + \lambda_2 - \mu_1 - \mu_2 - 8 = 0$

The phase 1 problem—reformulated

Minimize w, subject to:

$$
\begin{array}{ccccccccccccl}
-w & & -2x_2 & -2\mu_1 & -\mu_2 & +\lambda_1 & +\lambda_2 & & & & & = & -8 \\
 & 2x_1 & -2x_2 & +\mu_1 & -\mu_2 & -\lambda_1 & & & & +a_1 & & = & 2 \\
 & -2x_1 & +4x_2 & +\mu_1 & +2\mu_2 & & -\lambda_2 & & & & +a_2 & = & 6 \\
 & x_1 & +x_2 & & & & & +s_1 & & & & = & 2 \\
 & -x_1 & +2x_2 & & & & & & +s_2 & & & = & 2 \\
 & x_1, & x_2, & \mu_1, & \mu_2, & \lambda_1, & \lambda_2, & s_1, & s_2, & a_1, & a_2 & \geq & 0
\end{array}
$$

under the complementarity conditions:
$\mu_1 s_1 = \mu_2 s_2 = \lambda_1 x_1 = \lambda_2 x_2 = 0$

Solution of the Phase 1 problem on next page...

basis	w	x_1	x_2	μ_1	μ_2	λ_1	λ_2	s_1	s_2	a_1	a_2	RHS	
w	-1	0	-2	-2	-1	1	1	0	0	0	0	-8	x_2 in?
a_1	0	2	-2	1	-1	-1	0	0	0	1	0	2	$\lambda_2 = 0$
a_2	0	-2	4	1	2	0	-1	0	0	0	1	6	⇒ OK
s_1	0	1	1	0	0	0	0	1	0	0	0	2	s_2 out
s_2	0	-1	**2**	0	0	0	0	0	1	0	0	2	
w	-1	-1	0	-2	-1	1	1	0	1	0	0	-6	μ_1 in?
a_1	0	1	0	1	-1	-1	0	0	1	1	0	4	s_1 basic
a_2	0	0	0	1	2	0	-1	0	-2	0	1	2	⇒ no
s_1	0	**3/2**	0	0	0	0	0	1	-1/2	0	0	1	x_1 in?
x_2	0	-1/2	1	0	0	0	0	0	1/2	0	0	1	OK, s_1 out
w	-1	0	0	-2	-1	1	1	2/3	2/3	0	0	-16/3	μ_1 in?
a_1	0	0	0	1	-1	-1	0	-2/3	4/3	1	0	10/3	$s_1 = 0$
a_2	0	0	0	**1**	2	0	-1	0	-2	0	1	2	⇒ OK
x_1	0	1	0	0	0	0	0	2/3	-1/3	0	0	2/3	a_2 out
x_2	0	0	1	0	0	0	0	1/3	1/3	0	0	4/3	
w	-1	0	0	0	3	1	-1	2/3	-10/3	0	2	-4/3	s_2 in?
a_1	0	0	0	0	-3	-1	1	-2/3	**10/3**	1	-1	4/3	$\mu_2 = 0$
μ_1	0	0	0	1	2	0	-1	0	-2	0	1	2	⇒ OK
x_1	0	1	0	0	0	0	0	2/3	-1/3	0	0	2/3	a_1 out
x_2	0	0	1	0	0	0	0	1/3	1/3	0	0	4/3	
w	-1	0	0	0	0	0	0	0	0	1	1	0	optimum
s_2	0	0	0	0	-9/10	-3/10	3/10	-1/5	1	3/10	-3/10	2/5	
μ_1	0	0	0	1	1/5	-3/5	-2/5	-2/5	0	3/5	2/5	14/5	
x_1	0	1	0	0	-3/10	-1/10	1/10	3/5	0	1/10	-1/10	4/5	
x_2	0	0	1	0	3/10	1/10	-1/10	2/5	0	-1/10	1/10	6/5	

The optimal solution to the Phase 1 problem is given by:

$$\begin{bmatrix} x_1^* = 4/5, & x_2^* = 6/5 \\ \mu_1^* = 14/5, & \mu_2^* = 0 \\ \lambda_1^* = 0, & \lambda_2^* = 0 \\ s_1^* = 0, & s_2^* = 2/5 \end{bmatrix}$$

Note that:

$$\mu_1 s_1 = \mu_2 s_2 = \lambda_1 x_1 = \lambda_2 x_2 = 0$$

The original QP:

$$\begin{array}{rrcl} \text{minimize} & f(\mathbf{x}) & = & -2x_1 - 6x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{subject to} & & & x_1 + x_2 \leq 2 \\ & & & -x_1 + 2x_2 \leq 2 \\ & & & x_1 , x_2 \geq 0 \end{array}$$

$\Rightarrow f(\mathbf{x}^*) = -36/5$

What if f was not convex (i.e., $\mathbf{Q}$ not positive (semi)definite)?