
copyright (c) Nalisa Prints 2001

Some Mathematics for Radiation Therapy

- Adi Ben-Israel (Rutgers U.)

Some Mathematics for Radiation Therapy

- Adi Ben-Israel (Rutgers U.)

representing ideas of, and joint work with:

- Yair Censor (Haifa U.)
- Arik Hatwell (Haifa U.)
- Yuri Levin (Queens U.)

Some Mathematics for Radiation Therapy

• Adi Ben-Israel (Rutgers U.)

representing ideas of, and joint work with:

• Yair Censor (Haifa U.)
• Arik Hatwell (Haifa U.)
• Yuri Levin (Queens U.)

TJU Hospital, Phila, July 15, 2005

Thanks!

Plan

Plan

1. LP formulation of IMRT
   - duality
   - feasibility
   - relaxations.

Plan

1. LP formulation of IMRT
   - duality
   - feasibility
   - relaxations.
2. A Newton bracketing method for convex minimization
   - idea
   - illustration
   - advantages and disadvantages.

Plan

1. LP formulation of IMRT
   - duality
   - feasibility
   - relaxations.
2. A Newton bracketing method for convex minimization
   - idea
   - illustration
   - advantages and disadvantages.

If time permits:

3. Alternatives to quadratic objective,
4. Robust solution.

Intensity Modulated Radiation Therapy (IMRT)

- the (treatment) beams are indexed by j
- the intensity of the j_{th} beam is x_j
- the body voxels are indexed by i
- the radiation on the i_{th} voxel is modelled by

$$\sum_j a_{ij}\, x_j$$

- if the i_{th} voxel is in a target, the radiation has a lower bound

$$\sum_j a_{ij}\, x_j \geq \mathbf{l}_i$$

- if the i_{th} voxel is in an OAR, there is an upper bound

$$\sum_j a_{ij}\, x_j \leq \mathbf{u}_i$$

The IMRT (inverse) problem

Given: matrics A_1, A_2 and vectors $\mathbf{u}_1, \mathbf{l}_2$.

Find: a vector $\mathbf{x} = (x_j)$ such that

$$\begin{aligned} A_1\,\mathbf{x} &\leq \mathbf{u}_1 \\ \mathbf{l}_2 \leq A_2\,\mathbf{x} & \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{IMRT}$$

where:

- the subscript 1 refers to the OAR, and
- the subscript 2 to the target.

The IMRT (inverse) problem

Given: matrics A_1, A_2 and vectors $\mathbf{u}_1, \mathbf{l}_2$.

Find: a vector $\mathbf{x} = (x_j)$ such that

$$\begin{aligned} A_1\,\mathbf{x} &\le \mathbf{u}_1 \\ \mathbf{l}_2 \le A_2\,\mathbf{x} & \\ \mathbf{x} &\ge \mathbf{0} \end{aligned} \tag{IMRT}$$

where:

- the subscript 1 refers to the OAR, and
- the subscript 2 to the target.

This simplistic formulation will do for now.

Linear Programming (LP)

We write, WLOG, a LP problem as:

$$\begin{aligned} \min\ & \mathbf{c} \cdot \mathbf{x} \\ \text{s.t.} \quad & A\,\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{P}$$

the **primal problem**.

Linear Programming (LP)

We write, WLOG, a LP problem as:

$$\begin{aligned} &\min\ \mathbf{c}\cdot\mathbf{x} \\ &\text{s.t.}\quad A\,\mathbf{x}\geq\mathbf{b} \\ &\qquad\quad \mathbf{x}\geq\mathbf{0} \end{aligned} \tag{P}$$

the **primal problem**. The **dual problem** is

$$\begin{aligned} &\max\ \mathbf{b}\cdot\mathbf{y} \\ &\text{s.t.}\quad A^T\,\mathbf{y}\leq\mathbf{c} \\ &\qquad\quad \mathbf{y}\geq\mathbf{0} \end{aligned} \tag{D}$$

Linear Programming (LP)

We write, WLOG, a LP problem as:

$$\begin{aligned} \min\ & \mathbf{c}\cdot\mathbf{x} \\ \text{s.t.}\quad & A\,\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{P}$$

the **primal problem**. The **dual problem** is

$$\begin{aligned} \max\ & \mathbf{b}\cdot\mathbf{y} \\ \text{s.t.}\quad & A^T\,\mathbf{y} \leq \mathbf{c} \\ & \mathbf{y} \geq \mathbf{0} \end{aligned} \tag{D}$$

If $\mathbf{x}$ and $\mathbf{y}$ are feasible solutions of (P), (D) resp. then

$$\mathbf{c}\cdot\mathbf{x} \geq \mathbf{b}\cdot\mathbf{y}$$

Duality in LP

$$\begin{aligned} \min\ \mathbf{c}\cdot\mathbf{x} & & \quad (P) \\ \text{s.t. } A\,\mathbf{x} &\geq \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \qquad\qquad \begin{aligned} \max\ \mathbf{b}\cdot\mathbf{y} & & \quad (D) \\ \text{s.t. } A^T\,\mathbf{y} &\leq \mathbf{c} \\ \mathbf{y} &\geq \mathbf{0} \end{aligned}$$

These dual problems satisfy (exactly) one of the following 4 cases:

Duality in LP

$$\begin{array}{rclcrcl} \min \ \mathbf{c}\cdot\mathbf{x} & & & (P) & \max \ \mathbf{b}\cdot\mathbf{y} & & & (D) \\ \text{s.t. } A\,\mathbf{x} & \geq & \mathbf{b} & & \text{s.t. } A^T\,\mathbf{y} & \leq & \mathbf{c} \\ \mathbf{x} & \geq & \mathbf{0} & & \mathbf{y} & \geq & \mathbf{0} \end{array}$$

These dual problems satisfy (exactly) one of the following 4 cases:

1. (P) feasible, (D) feasible, min (P) = max (D)

Duality in LP

$$\begin{array}{rcl} \min\ \mathbf{c}\cdot\mathbf{x} & & (P) \\ \text{s.t. } A\,\mathbf{x} & \geq & \mathbf{b} \\ \mathbf{x} & \geq & \mathbf{0} \end{array} \qquad \begin{array}{rcl} \max\ \mathbf{b}\cdot\mathbf{y} & & (D) \\ \text{s.t. } A^T\,\mathbf{y} & \leq & \mathbf{c} \\ \mathbf{y} & \geq & \mathbf{0} \end{array}$$

These dual problems satisfy (exactly) one of the following 4 cases:

1. (P) feasible, (D) feasible, min (P) = max (D)

2. (P) feasible and unbounded, (D) infeasible
   (P) unbounded means: $\exists$ feasible $\mathbf{x}_k$ s.t. $\mathbf{c}\cdot\mathbf{x}_k \to -\infty$

Duality in LP

$$\begin{aligned} \min\ \mathbf{c}\cdot\mathbf{x} & & \text{(P)} \\ \text{s.t. } A\,\mathbf{x} &\geq \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \qquad \begin{aligned} \max\ \mathbf{b}\cdot\mathbf{y} & & \text{(D)} \\ \text{s.t. } A^T\mathbf{y} &\leq \mathbf{c} \\ \mathbf{y} &\geq \mathbf{0} \end{aligned}$$

These dual problems satisfy (exactly) one of the following 4 cases:

1. (P) feasible, (D) feasible, min (P) = max (D)
2. (P) feasible and unbounded, (D) infeasible
   (P) unbounded means: $\exists$ feasible $\mathbf{x}_k$ s.t. $\mathbf{c}\cdot\mathbf{x}_k \to -\infty$
3. (P) infeasible, (D) unbounded
   (D) unbounded means: $\exists$ feasible $\mathbf{y}_k$ s.t. $\mathbf{b}\cdot\mathbf{y}_k \to \infty$

Duality in LP

$$\begin{array}{rcl} \min \ \mathbf{c}\cdot\mathbf{x} & & \text{(P)} \\ \text{s.t. } A\,\mathbf{x} & \geq & \mathbf{b} \\ \mathbf{x} & \geq & \mathbf{0} \end{array} \qquad \begin{array}{rcl} \max \ \mathbf{b}\cdot\mathbf{y} & & \text{(D)} \\ \text{s.t. } A^T\,\mathbf{y} & \leq & \mathbf{c} \\ \mathbf{y} & \geq & \mathbf{0} \end{array}$$

These dual problems satisfy (exactly) one of the following 4 cases:

1. (P) feasible, (D) feasible, min (P) = max (D)
2. (P) feasible and unbounded, (D) infeasible
   (P) unbounded means: $\exists$ feasible $\mathbf{x}_k$ s.t. $\mathbf{c}\cdot\mathbf{x}_k \to -\infty$
3. (P) infeasible, (D) unbounded
   (D) unbounded means: $\exists$ feasible $\mathbf{y}_k$ s.t. $\mathbf{b}\cdot\mathbf{y}_k \to \infty$
4. (P) and (D) infeasible.

Interpretation

Consider the **perturbed problems**

$$
\begin{array}{rcl|rcl}
p(\mathbf{u}) := \min \ \mathbf{c}\cdot\mathbf{x} & & (\mathrm{P}.\mathbf{u}) & d(\mathbf{v}) := \max \ \mathbf{b}\cdot\mathbf{y} & & (\mathrm{D}.\mathbf{v}) \\
\text{s.t. } A\,\mathbf{x} & \geq & \mathbf{b}+\mathbf{u} & \text{s.t. } A^T\,\mathbf{y} & \leq & \mathbf{c}+\mathbf{v} \\
\mathbf{x} & \geq & \mathbf{0} & \mathbf{y} & \geq & \mathbf{0}
\end{array}
$$

where (P) = (P.$\mathbf{0}$), (D) = (D.$\mathbf{0}$). Then:

Interpretation

Consider the **perturbed problems**

$\begin{aligned} p(\mathbf{u}) := \min\ & \mathbf{c}\cdot\mathbf{x} & & \text{(P.}\mathbf{u}\text{)} \\ \text{s.t. } A\,\mathbf{x} & \geq & & \mathbf{b}+\mathbf{u} \\ \mathbf{x} & \geq & & \mathbf{0} \end{aligned}$	$\begin{aligned} d(\mathbf{v}) := \max\ & \mathbf{b}\cdot\mathbf{y} & & \text{(D.}\mathbf{v}\text{)} \\ \text{s.t. } A^T\,\mathbf{y} & \leq & & \mathbf{c}+\mathbf{v} \\ \mathbf{y} & \geq & & \mathbf{0} \end{aligned}$

where (P) = (P.$\mathbf{0}$), (D) = (D.$\mathbf{0}$). Then:

- $p(\cdot)$ is convex
- $d(\cdot)$ is concave

Interpretation

Consider the **perturbed problems**

$$p(\mathbf{u}) := \min \mathbf{c}\cdot\mathbf{x} \quad (\mathrm{P}.\mathbf{u}) \qquad \text{s.t. } A\mathbf{x} \ge \mathbf{b}+\mathbf{u}, \quad \mathbf{x} \ge \mathbf{0}$$

$$d(\mathbf{v}) := \max \mathbf{b}\cdot\mathbf{y} \quad (\mathrm{D}.\mathbf{v}) \qquad \text{s.t. } A^T\mathbf{y} \le \mathbf{c}+\mathbf{v}, \quad \mathbf{y} \ge \mathbf{0}$$

where (P) = (P.$\mathbf{0}$), (D) = (D.$\mathbf{0}$). Then:

- $p(\cdot)$ is convex
- $d(\cdot)$ is concave
- if $\mathbf{y}^*$ is an optimal solution of (D) then $\mathbf{y}^* \in \partial p(\mathbf{0})$

  $\therefore$ if $p(\cdot)$ is differentiable at $\mathbf{0}$ then $\mathbf{y}^* = \nabla p(\mathbf{0})$

Interpretation

Consider the **perturbed problems**

$$p(\mathbf{u}) := \min \ \mathbf{c}\cdot\mathbf{x} \quad (\text{P}.\mathbf{u}) \qquad \text{s.t. } A\mathbf{x} \geq \mathbf{b}+\mathbf{u}, \quad \mathbf{x} \geq \mathbf{0}$$

$$d(\mathbf{v}) := \max \ \mathbf{b}\cdot\mathbf{y} \quad (\text{D}.\mathbf{v}) \qquad \text{s.t. } A^T\mathbf{y} \leq \mathbf{c}+\mathbf{v}, \quad \mathbf{y} \geq \mathbf{0}$$

where (P) = (P.$\mathbf{0}$), (D) = (D.$\mathbf{0}$). Then:

- $p(\cdot)$ is convex
- $d(\cdot)$ is concave
- if $\mathbf{y}^*$ is an optimal solution of (D) then $\mathbf{y}^* \in \partial p(\mathbf{0})$

  $\therefore$ if $p(\cdot)$ is differentiable at $\mathbf{0}$ then $\mathbf{y}^* = \nabla p(\mathbf{0})$
- if $\mathbf{x}^*$ is an optimal solution of (P) then $\mathbf{x}^* \in \partial d(\mathbf{0})$

  $\therefore$ if $d(\cdot)$ is differentiable at $\mathbf{0}$ then $\mathbf{x}^* = \nabla d(\mathbf{0})$

LP formulation of IMRT

The IMRT problem is: find vector $\mathbf{x}$ such that

$$A_1\mathbf{x} \le \mathbf{u}_1, \quad \mathbf{l}_2 \le A_2\mathbf{x}, \quad \mathbf{x} \ge \mathbf{0} \qquad \text{(IMRT)}$$

LP formulation of IMRT

The IMRT problem is: find vector $\mathbf{x}$ such that

$$A_1\mathbf{x} \le \mathbf{u}_1, \quad \mathbf{l}_2 \le A_2\mathbf{x}, \quad \mathbf{x} \ge \mathbf{0} \qquad \text{(IMRT)}$$

This problem is equivalent to the LP

$$\begin{aligned} \min \quad & \mathbf{0} \cdot \mathbf{x} & & \text{(P)} \\ \text{s.t.} \quad & -A_1\mathbf{x} \ge -\mathbf{u}_1 & (\mathbf{y}_1) & \\ & A_2\mathbf{x} \ge \mathbf{l}_2 & (\mathbf{y}_2) & \\ & \mathbf{x} \ge \mathbf{0} & & \end{aligned}$$

with dual variables $\mathbf{y}_1, \mathbf{y}_2$.

Duality

We saw that IMRT is equivalent to the LP

$$
\begin{aligned}
\min \quad & \mathbf{0}\cdot\mathbf{x} & & \text{(P)}\\
\text{s.t.} \quad & -A_1\mathbf{x} \geq -\mathbf{u}_1 & (\mathbf{y}_1) &\\
& A_2\mathbf{x} \geq \mathbf{l}_2 & (\mathbf{y}_2) &\\
& \mathbf{x} \geq \mathbf{0} & &
\end{aligned}
$$

with dual variables $\mathbf{y}_1, \mathbf{y}_2$.

Duality

We saw that IMRT is equivalent to the LP

$$\begin{aligned} \min \quad & \mathbf{0} \cdot \mathbf{x} & & \text{(P)} \\ \text{s.t.} \quad & -A_1 \mathbf{x} \geq -\mathbf{u}_1 & (\mathbf{y}_1) & \\ & A_2 \mathbf{x} \geq \mathbf{l}_2 & (\mathbf{y}_2) & \\ & \mathbf{x} \geq \mathbf{0} & & \end{aligned}$$

with dual variables $\mathbf{y}_1, \mathbf{y}_2$. The dual problem is

$$\begin{aligned} \max \quad & -\mathbf{u}_1 \cdot \mathbf{y}_1 + \mathbf{l}_2 \cdot \mathbf{y}_2 & \text{(D)} \\ \text{s.t.} \quad & -A_1^T \mathbf{y}_1 + A_2^T \mathbf{y}_2 \leq \mathbf{0} & \\ & \mathbf{y}_1, \mathbf{y}_2 \geq \mathbf{0} & \end{aligned}$$

Duality

We saw that IMRT is equivalent to the LP

$$\begin{aligned} \min \quad & \mathbf{0} \cdot \mathbf{x} & & \text{(P)} \\ \text{s.t.} \quad & -A_1 \mathbf{x} \geq -\mathbf{u}_1 & (\mathbf{y}_1) & \\ & A_2 \mathbf{x} \geq \mathbf{l}_2 & (\mathbf{y}_2) & \\ & \mathbf{x} \geq \mathbf{0} & & \end{aligned}$$

with dual variables $\mathbf{y}_1, \mathbf{y}_2$. The dual problem is

$$\begin{aligned} \max \quad & -\mathbf{u}_1 \cdot \mathbf{y}_1 + \mathbf{l}_2 \cdot \mathbf{y}_2 & \text{(D)} \\ \text{s.t.} \quad & -A_1^T \mathbf{y}_1 + A_2^T \mathbf{y}_2 \leq \mathbf{0} & \\ & \mathbf{y}_1, \mathbf{y}_2 \geq \mathbf{0} & \end{aligned}$$

The dual is (trivially) feasible.

Relaxation of IMRT

If (IMRT) is infeasible, then so is (P), and (D) is unbounded. In this case we can relax (IMRT) by replacing the constraints

$$A_1 \mathbf{x} \leq \mathbf{u}_1$$

by

$$A_1 \mathbf{x} \leq U_1 \, \mathbf{t}$$

where U_1 is a diagonal matrix with the vector $\mathbf{u}_1$ on its diagonal, and the vector $\mathbf{t}$ satisfies

$$\mathbf{0} \leq \mathbf{t} \leq (1+\beta) \, \mathbf{e}$$

Relaxation of IMRT

If (IMRT) is infeasible, then so is (P), and (D) is unbounded. In this case we can relax (IMRT) by replacing the constraints

$$A_1 \mathbf{x} \leq \mathbf{u}_1$$

by

$$A_1 \mathbf{x} \leq U_1 \, \mathbf{t}$$

where U_1 is a diagonal matrix with the vector $\mathbf{u}_1$ on its diagonal, and the vector $\mathbf{t}$ satisfies

$$\mathbf{0} \leq \mathbf{t} \leq (1+\beta) \, \mathbf{e}$$

The parameter β is a measure of the relaxation (violation) of the constraints.

Relaxation of (P)

A corresponding relaxation of (P) is the problem

$$
\begin{aligned}
\min \quad & \mathbf{0}\cdot\mathbf{t} + \mathbf{0}\cdot\mathbf{x} & & \text{(RP)}\\
\text{s.t.} \quad & U_1\mathbf{t} - A_1\mathbf{x} \geq \mathbf{0} & (\mathbf{y}_1) &\\
& A_2\mathbf{x} \geq \mathbf{l}_2 & (\mathbf{y}_2) &\\
& -\mathbf{t} \geq -(1+\beta)\,\mathbf{e} & (\mathbf{z}) &\\
& \mathbf{t}, \mathbf{x} \geq \mathbf{0} & &
\end{aligned}
$$

Relaxation of (P)

A corresponding relaxation of (P) is the problem

$$
\begin{array}{rll}
\min & \mathbf{0}\cdot\mathbf{t}+\mathbf{0}\cdot\mathbf{x} & \text{(RP)} \\
\text{s.t.} & U_1\mathbf{t}-A_1\mathbf{x}\geq\mathbf{0} & (\mathbf{y}_1) \\
& A_2\mathbf{x}\geq\mathbf{l}_2 & (\mathbf{y}_2) \\
& -\mathbf{t}\geq-(1+\beta)\,\mathbf{e} & (\mathbf{z}) \\
& \mathbf{t},\mathbf{x}\geq\mathbf{0} &
\end{array}
$$

The dual of (RP) is

$$
\begin{array}{rll}
\max & \mathbf{l}_2\cdot\mathbf{y}_2-(1+\beta)\,\mathbf{e}\cdot\mathbf{z} & \text{(RD)} \\
\text{s.t.} & -A_1^T\mathbf{y}_1+A_2^T\mathbf{y}_2\leq\mathbf{0} & \\
& U_1\mathbf{y}_1-\mathbf{z}\leq\mathbf{0} & \\
& \mathbf{y}_1,\mathbf{y}_2,\mathbf{z}\geq\mathbf{0} &
\end{array}
$$

Relaxation may not work in practice

The optimal $\mathbf{z}$ in (RD) satisfies

$$\mathbf{z} = U_1 \mathbf{y}_1 \tag{1}$$

and (RD) reduces to

$$\begin{aligned} \max \quad & -(1+\beta)\, \mathbf{u}_1 \cdot \mathbf{y}_1 + \mathbf{l}_2 \cdot \mathbf{y}_2 \\ \text{s.t.} \quad & -A_1^T \mathbf{y}_1 + A_2^T \mathbf{y}_2 \le \mathbf{0} \\ & \mathbf{y}_1, \mathbf{y}_2 \ge \mathbf{0} \end{aligned} \tag{RD*}$$

Relaxation may not work in practice

The optimal $\mathbf{z}$ in (RD) satisfies

$$\mathbf{z} = U_1 \mathbf{y}_1 \tag{1}$$

and (RD) reduces to

$$\begin{aligned} \max \ & -(1+\beta)\, \mathbf{u}_1 \cdot \mathbf{y}_1 + \mathbf{l}_2 \cdot \mathbf{y}_2 \\ \text{s.t.} \quad & - A_1^T \mathbf{y}_1 + A_2^T \mathbf{y}_2 \leq \mathbf{0} \\ & \mathbf{y}_1, \mathbf{y}_2 \geq \mathbf{0} \end{aligned} \tag{RD*}$$

If (D) is unbounded, then – for some sufficiently large $\beta > 0$ – (RD*) will be bounded, and the relaxed (IMRT) feasible.

Relaxation may not work in practice

The optimal $\mathbf{z}$ in (RD) satisfies

$$\mathbf{z} = U_1 \mathbf{y}_1 \tag{1}$$

and (RD) reduces to

$$\begin{aligned} \max \ & -(1+\beta)\, \mathbf{u}_1 \cdot \mathbf{y}_1 + \mathbf{l}_2 \cdot \mathbf{y}_2 \\ \text{s.t.} \quad & -A_1^T \mathbf{y}_1 + A_2^T \mathbf{y}_2 \leq \mathbf{0} \\ & \mathbf{y}_1, \mathbf{y}_2 \geq \mathbf{0} \end{aligned} \tag{RD*}$$

If (D) is unbounded, then – for some sufficiently large $\beta > 0$ – (RD*) will be bounded, and the relaxed (IMRT) feasible.

However, the smallest such β may be too large.

Implementation issues

- In practice, there are several OAR's with different relaxations.
- An $\{\alpha, \beta\}$ relaxation is when at most α % of the constraints are relaxed by β, i.e. replaced by $\sum_j a_{ij}\, x_j \leq u_i\, t_i, \quad 0 \leq t_i \leq 1 + \beta$

Implementation issues

- In practice, there are several OAR's with different relaxations.
- An $\{\alpha, \beta\}$ relaxation is when at most α % of the constraints are relaxed by β, i.e. replaced by $\sum_j a_{ij}\, x_j \le u_i\, t_i, \quad 0 \le t_i \le 1 + \beta$
- in (RP) above we can replace the objective by

$$\min \sum_i t_i \;, \quad \text{minimizing the sum of the violations.}$$

Implementation issues

- In practice, there are several OAR's with different relaxations.
- An $\{\alpha, \beta\}$ relaxation is when at most α % of the constraints are relaxed by β, i.e. replaced by $\sum_j a_{ij}\, x_j \le u_i\, t_i, \quad 0 \le t_i \le 1 + \beta$
- in (RP) above we can replace the objective by

$$\min \sum_i t_i \ , \quad \text{minimizing the sum of the violations.}$$

- LP uses extreme point solutions, so a good many of the relaxed constraints above will be at their limit $t_i = 1 + \beta$.
- A rough check of $\{\alpha, \beta\}$ relaxation of N_1 inequalities is

$$\sum_i t_i \le N_1(1 + \alpha\beta)$$

Convex minimization in IMRT

Solving the IMRT problem often requires the (unconstrained) minimization of a convex function.

Convex minimization in IMRT

Solving the IMRT problem often requires the (unconstrained) minimization of a convex function.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if minorized by its tangents, i.e. if the tangent to its graph at any point $\mathbf{x}$ is below the graph. If f is differentiable this means

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y}$$

Convex minimization in IMRT

Solving the IMRT problem often requires the (unconstrained) minimization of a convex function.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if minorized by its tangents, i.e. if the tangent to its graph at any point $\mathbf{x}$ is below the graph. If f is differentiable this means

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y}$$

A **quadratic (form)** is a function

$$f(\mathbf{x}) = \mathbf{x} \cdot Q\mathbf{x} - \mathbf{c} \cdot \mathbf{x} + \gamma$$

where Q is a symmetric matrix. It is convex [strictly convex] if Q is positive semi-definite [positive definite], i.e. if all its eigenvalues are non-negative [positive].

The Newton Bracketing (NB) method for minimization of convex functions $f : \mathbb{R} \to \mathbb{R}$

Given: $f : \mathbb{R} \to \mathbb{R}$, convex, bounded below, its infimum attained.

The problem:

$$\min_{x \in \mathbb{R}} f(x) \;:\longrightarrow\; f_{\min}$$

The Newton Bracketing (NB) method for minimization of convex functions $f : \mathbb{R} \to \mathbb{R}$

Given: $f : \mathbb{R} \to \mathbb{R}$, convex, bounded below, its infimum attained.
The problem:

$$\min_{x \in \mathbb{R}} f(x) \;:\longrightarrow\; f_{\min}$$

- If $f_{\min}$ is known, solve

$$f(x) = f_{\min} \; .$$

The Newton Bracketing (NB) method for minimization of convex functions $f : \mathbb{R} \to \mathbb{R}$

Given: $f : \mathbb{R} \to \mathbb{R}$, convex, bounded below, its infimum attained.
The problem:

$$\min_{x \in \mathbb{R}} f(x) \;:\longrightarrow\; f_{\min}$$

- If $f_{\min}$ is known, solve

$$f(x) = f_{\min} \;.$$

- Otherwise, we generate a sequence of problems

$$f(x) = M^k \;, \quad \text{where } M^k \to f_{\min}$$

and do one Newton iteration per problem.

The Newton Bracketing (NB) method for minimization of convex functions $f : \mathbb{R} \to \mathbb{R}$

Given: $f : \mathbb{R} \to \mathbb{R}$, convex, bounded below, its infimum attained.
The problem:

$$\min_{x \in \mathbb{R}} f(x) \;:\longrightarrow\; f_{\min}$$

• If $f_{\min}$ is known, solve

$$f(x) = f_{\min} \; .$$

• Otherwise, we generate a sequence of problems

$$f(x) = M^k \; , \quad \text{where } M^k \to f_{\min}$$

and do one Newton iteration per problem.
• In each iteration a bracket $[L_k, U_k]$ containing $f_{\min}$ is reduced, so that $(U_k - L_k) \to 0$.

The NB method for minimizing f convex

Initially given:

- a lower bound L on $f_{\min}$
- a tolerance ϵ
- an initial solution x (arbitrary)

The NB method for minimizing f convex

Initially given:

- a lower bound L on $f_{\min}$
- a tolerance ϵ
- an initial solution x (arbitrary)

An **iteration** begins with:

- a lower bound L on $f_{\min}$
- a current solution x, and
- an upper bound $U := f(x)$ on $f_{\min}$.

The NB method for minimizing f convex

Initially given:

- a lower bound L on $f_{\min}$
- a tolerance ϵ
- an initial solution x (arbitrary)

An **iteration** begins with:

- a lower bound L on $f_{\min}$
- a current solution x, and
- an upper bound $U := f(x)$ on $f_{\min}$.

(1) **if** $U - L < \epsilon$ **stop**

(2) **else** **select** $L < M < U$

(3) **do** $x_+ := x - \dfrac{f(x) - M}{f'(x)}$

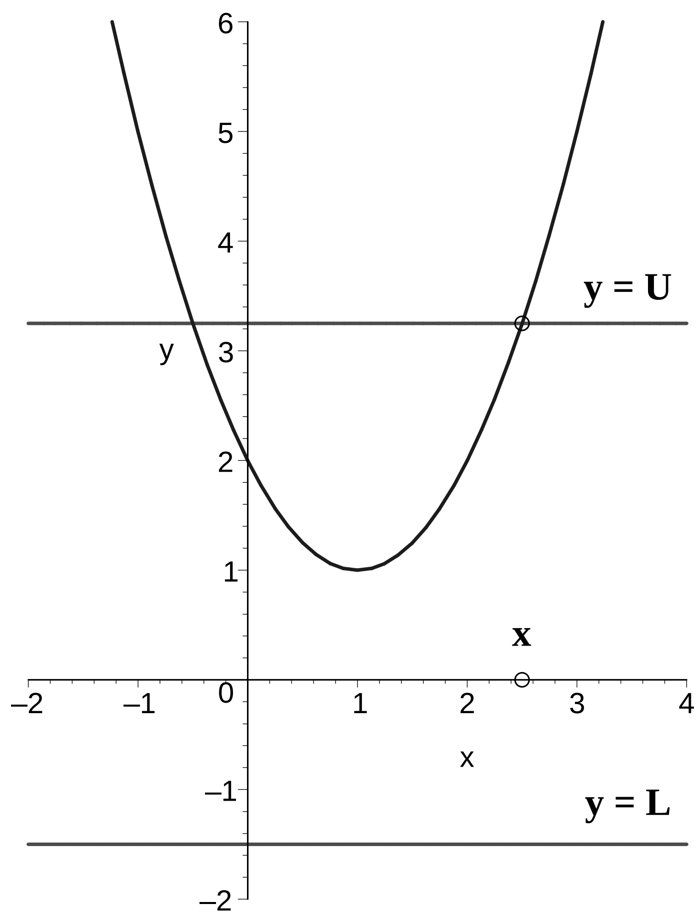
(4) **if** $f(x_+) < U$ **then** $U_+ := f(x_+)$

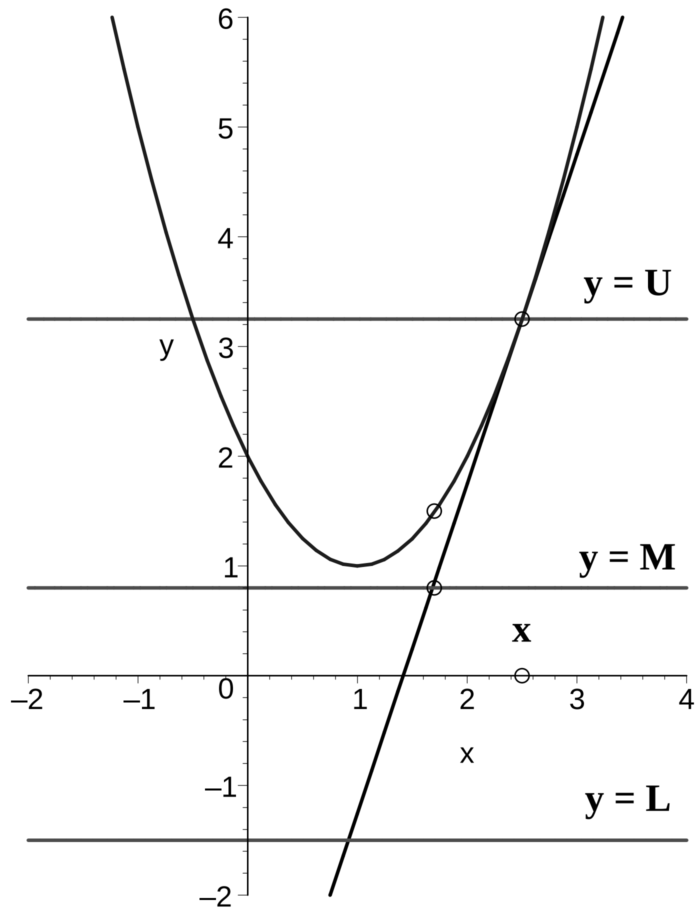
(5) **else** $L_+ := M,$ **return**

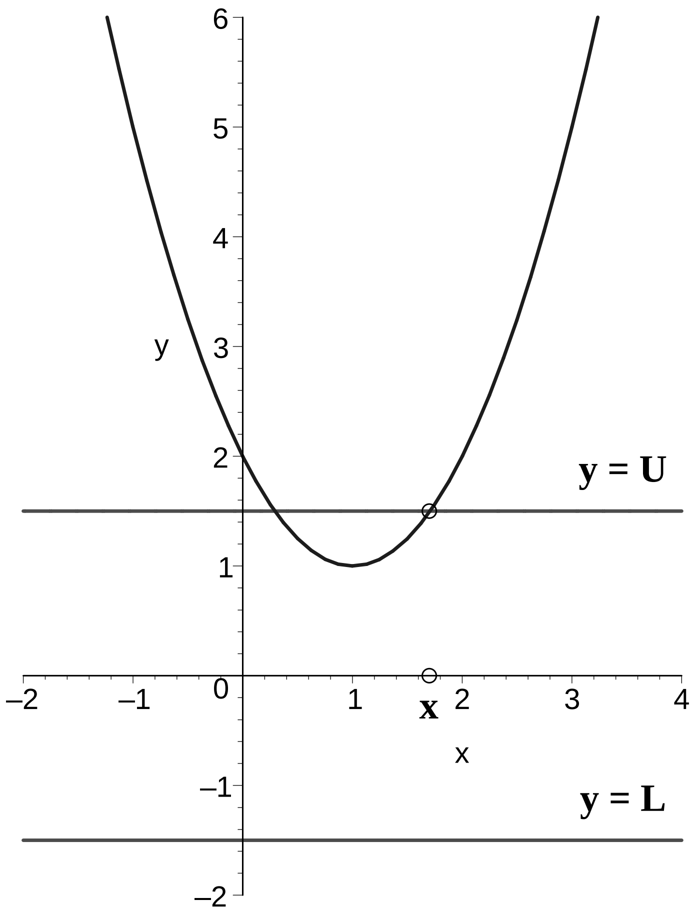
Illustration

The following slides show several iterations of the NB method.

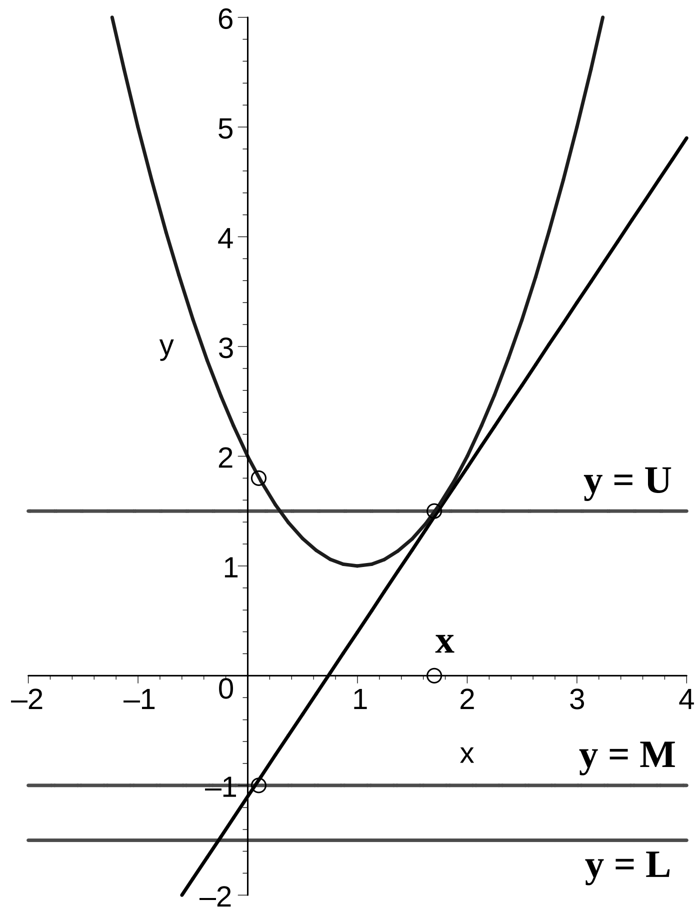
Note how the lines $y = U$ and $y = L$ come closer, this illustrates the shrinking of the bracket $U - L$.


6
5
4
y = U
y
3
2
1
x
0
–2
–1
1
2
3
4
x
–1
y = L
–2

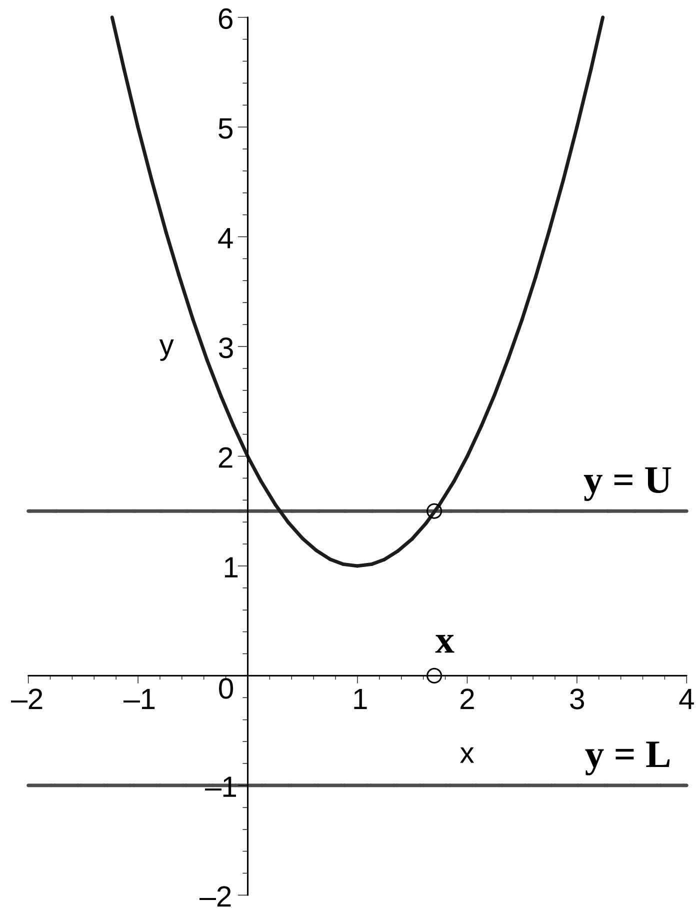

y = U
y = M
x
y = L
y
x
6
5
4
3
2
1
0
–1
–2
–2
–1
1
2
3
4

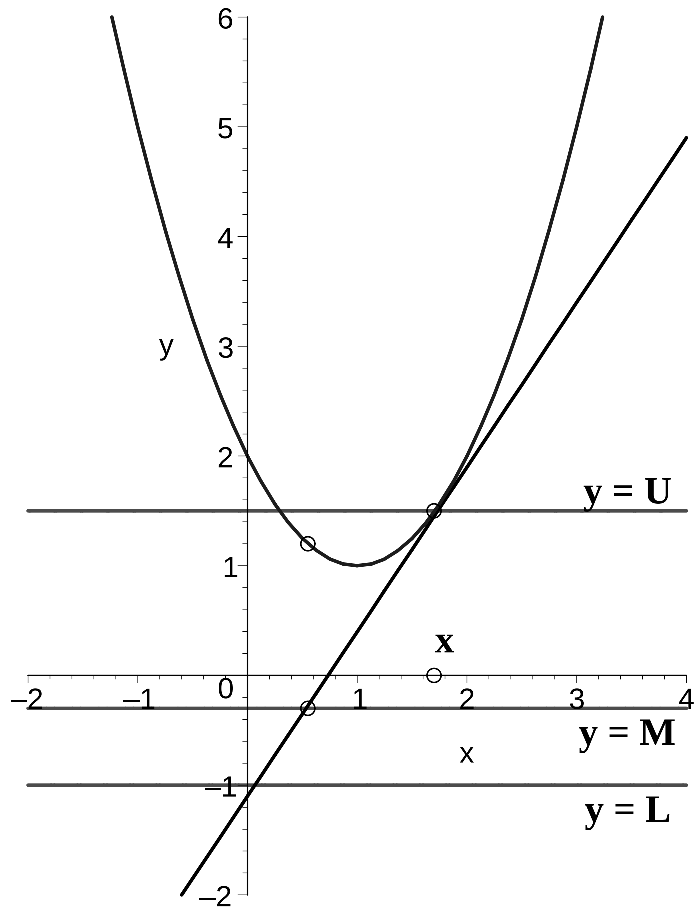


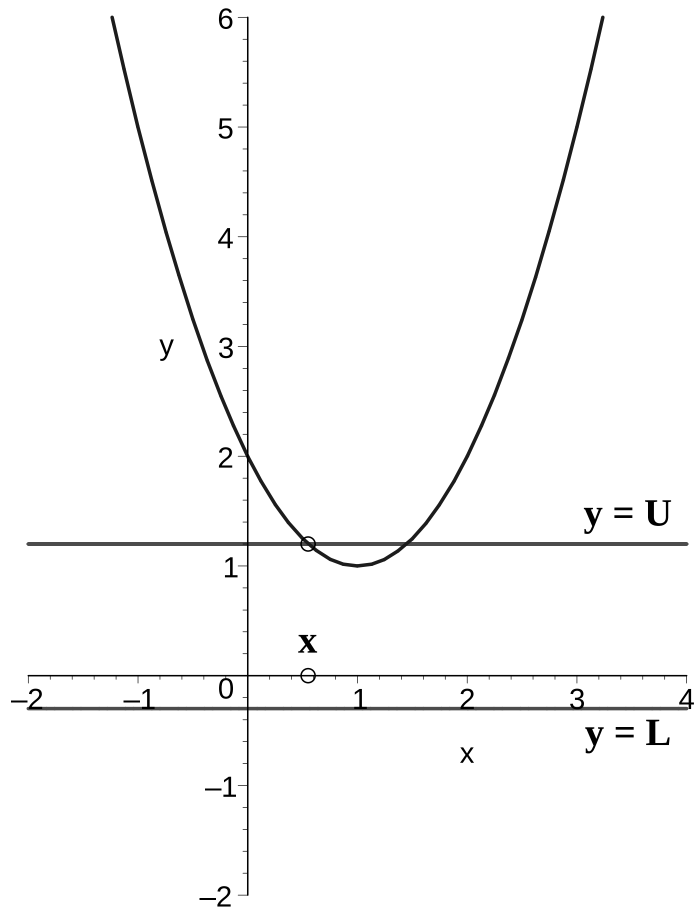
y
6
5
4
3
2
1
0
–1
–2
–2
–1
1
2
3
4
x
y = U
x
y = L

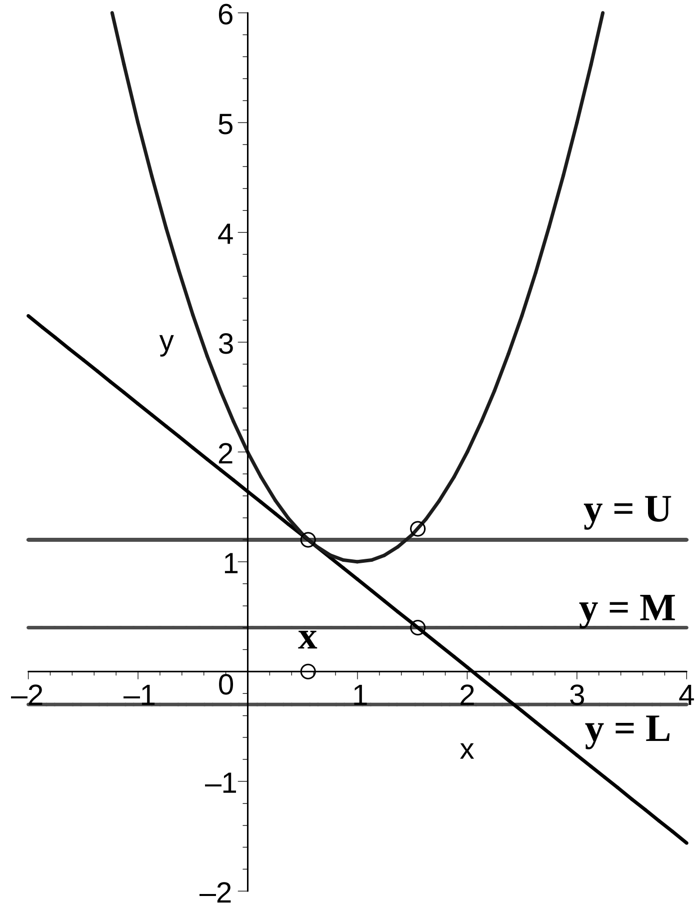


6
5
4
y
3
2
y = U
1
x
0
–2
–1
1
2
3
4
x
y = M
–1
y = L
–2

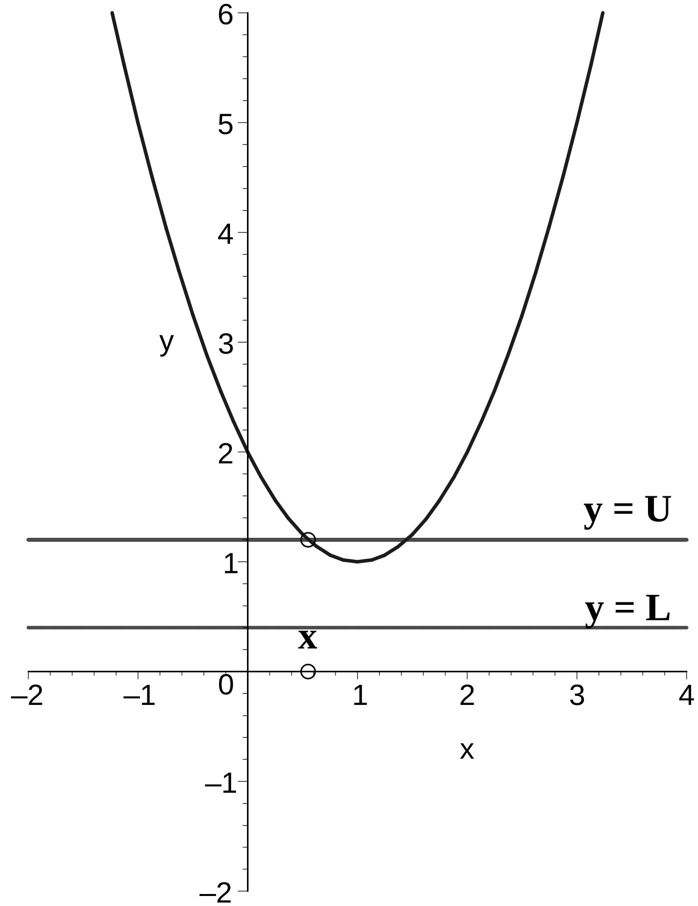

6
5
4
y
3
2
y = U
1
x
0
–2
–1
1
2
3
4
x
y = L
–1
–2


y = U
y = M
y = L
x
y
x
6
5
4
3
2
1
0
–1
–2
–2
–1
1
2
3
4

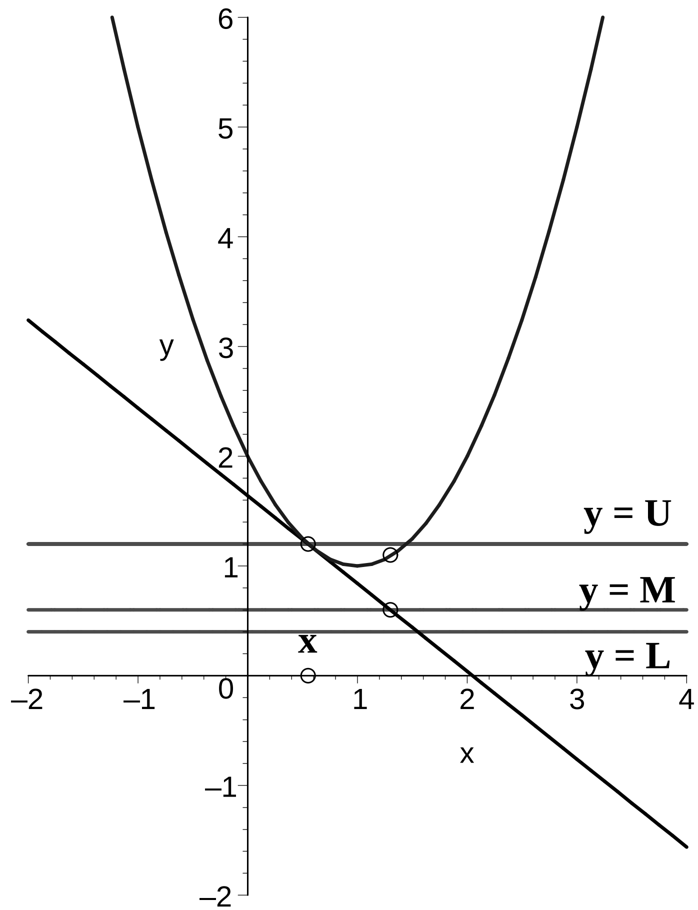

6
5
4
y
3
2
y = U
1
x
0
–2
–1
1
2
3
4
y = L
x
–1
–2

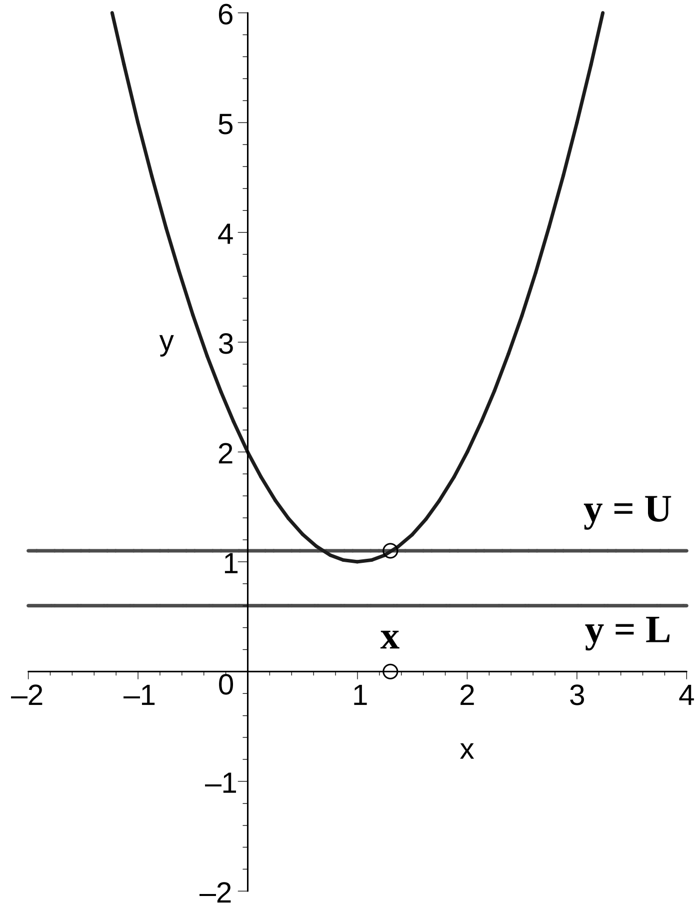


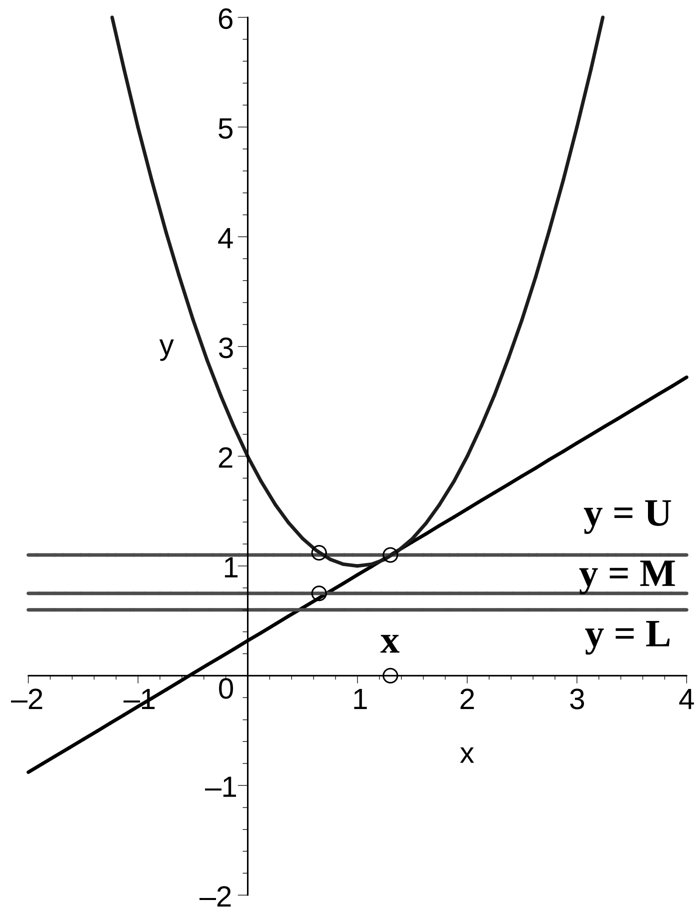
6
5
4
3
2
1
0
–1
–2
y
–2
–1
1
2
3
4
x
y = U
y = M
y = L
x



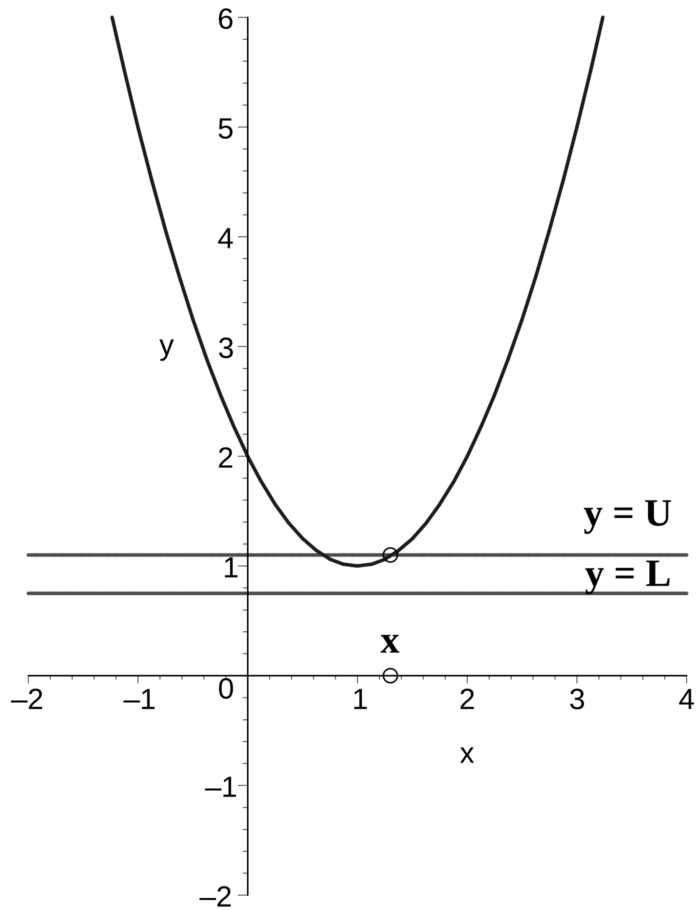
6
5
4
y
3
2
y = U
1
y = L
x
0
–2
–1
1
2
3
4
x
–1
–2


y = U
y = M
y = L
x
y
x


6
5
4
y
3
2
y = U
1
y = L
x
0
–2
–1
1
2
3
4
x
–1
–2



y = U
y = M
y = L
x
y
x



6
5
4
y
3
2
y = U
1
y = L
x
0
–2
–1
1
2
3
4
x
–1
–2

The NB iteration

(1)	**if**	$U - L < \epsilon$ **stop**
(2)	**else**	**select** $L < M < U$
(3)		**do** $x_+ := x - \dfrac{f(x) - M}{f'(x)}$
(4)	**if**	$f(x_+) < U$ **then** $U_+ := f(x_+)$
(5)	**else**	$L_+ := M,$ **return**

The NB iteration

(1) **if** $U - L < \epsilon$ **stop**

(2) **else** **select** $L < M < U$

(3) **do** $x_+ := x - \dfrac{f(x) - M}{f'(x)}$

(4) **if** $f(x_+) < U$ **then** $U_+ := f(x_+)$

(5) **else** $L_+ := M,$ **return**

- (1) is a stopping rule, compare with gradient methods.

The NB iteration

(1) **if** $U - L < \epsilon$ **stop**

(2) **else** **select** $L < M < U$

(3) **do** $x_+ := x - \dfrac{f(x) - M}{f'(x)}$

(4) **if** $f(x_+) < U$ **then** $U_+ := f(x_+)$

(5) **else** $L_+ := M,$ **return**

- (1) is a stopping rule, compare with gradient methods.
- In (2), $M = \alpha U + (1 - \alpha)L$ for some $0 < \alpha < 1$.

The NB iteration

(1)	**if**	$U - L < \epsilon$	**stop**
(2)	**else**	**select** $L < M < U$	
(3)		**do** $x_+ := x - \dfrac{f(x) - M}{f'(x)}$	
(4)	**if**	$f(x_+) < U$ **then** $U_+ := f(x_+)$	
(5)	**else**	$L_+ := M,$ **return**	

- (1) is a stopping rule, compare with gradient methods.
- In (2), $M = \alpha U + (1 - \alpha)L$ for some $0 < \alpha < 1$.
- In (5), $\mathbf{x}_+ := \mathbf{x}$, i.e. the solution is unchanged, only L is raised.

The NB iteration

$$
\begin{array}{llll}
(1) & \textbf{if} & U - L < \epsilon & \textbf{stop} \\
(2) & \textbf{else} & \textbf{select} & L < M < U \\
(3) & & \textbf{do} & x_+ := x - \dfrac{f(x) - M}{f'(x)} \\
(4) & \textbf{if} & f(x_+) < U & \textbf{then} \quad U_+ := f(x_+) \\
(5) & \textbf{else} & L_+ := M, & \textbf{return}
\end{array}
$$

- (1) is a stopping rule, compare with gradient methods.
- In (2), $M = \alpha U + (1 - \alpha)L$ for some $0 < \alpha < 1$.
- In (5), $\mathbf{x}_+ := \mathbf{x}$, i.e. the solution is unchanged, only L is raised.
- The bracket reduction is

$$
\frac{U_+ - L_+}{U - L} = \begin{cases} \dfrac{f(x_+) - L}{f(x) - L}, & \text{in (4);} \\ 1 - \alpha, & \text{in (5).} \end{cases}
$$

Directional Newton iteration for $f(\mathbf{x}) = 0,\ f : \mathbb{R}^n \to \mathbb{R}$

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\, \mathbf{d}$$

$\mathbf{d}$ a direction vector, $\|\mathbf{d}\| = 1$.

Directional Newton iteration for $f(\mathbf{x}) = 0,\ f : \mathbb{R}^n \to \mathbb{R}$

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\, \mathbf{d}$$

$\mathbf{d}$ a direction vector, $\|\mathbf{d}\| = 1$.

This corresponds to the Newton step for

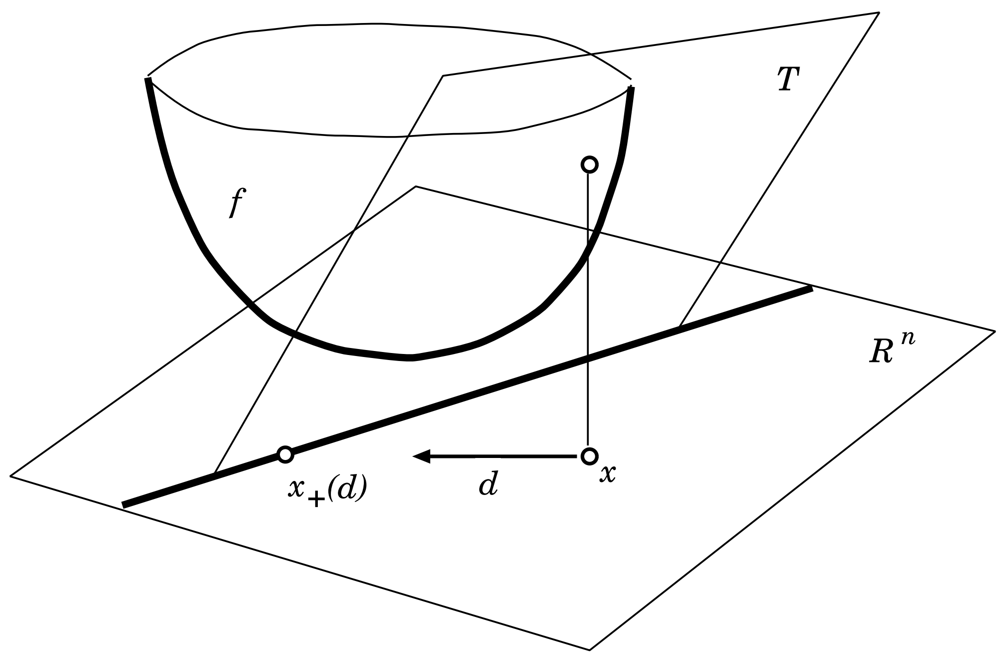
$$F(t) = f(\mathbf{x} + t\mathbf{d}),$$

at $t = 0$, using $F'(0) = f'(\mathbf{x}, \mathbf{d})$,

$$t_+ := -\frac{F(0)}{F'(0)}\ .$$

Geometric interpretation

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\, \mathbf{d}$$



The set $\{\mathbf{x}_+(\mathbf{d}) : \mathbf{d} \in \mathbb{R}^n\}$ is the intersection of $\mathbb{R}^n$ and the tangent hyperplane T (in $\mathbb{R}^{n+1}$) of the graph of f at $(\mathbf{x},\, f(\mathbf{x}))$.

The gradient direction

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\,\mathbf{d}$$

The gradient direction

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\,\mathbf{d}$$

The gradient direction

$$\begin{aligned} \mathbf{d} \quad &:= \quad \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|} \quad \text{gives} \\ \mathbf{x}_+ \quad &:= \quad \mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\,\nabla f(\mathbf{x})\ . \end{aligned}$$

The gradient direction

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\,\mathbf{d}$$

The gradient direction

$$\begin{aligned} \mathbf{d} &:= \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|} \quad \text{gives} \\ \mathbf{x}_+ &:= \mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\,\nabla f(\mathbf{x}) \,. \end{aligned}$$

For $n = 1$, this gives

$$x_+ := x - \frac{f(x)}{f'(x)}$$

the scalar Newton method.

The NB method for convex functions: $\mathbb{R}^n \to \mathbb{R}$

Iteration begins with a current solution $\mathbf{x}$, and bounds
$L < f_{\min} \le U \;,\; U := f(\mathbf{x})$

1	**if**	$U - L < \epsilon$	**solution** $:= \mathbf{x}$, **stop**
2	**endif**	**select**	$M := \alpha\, U + (1-\alpha)\, L$
3		**do**	$\mathbf{x}_+ := \mathbf{x} - \dfrac{f(\mathbf{x}) - M}{\lVert\nabla f\rVert^2}\,\nabla f(\mathbf{x})$
4	**if**	$f(\mathbf{x}_+) < f(\mathbf{x})$	**set** $U_+ := f(\mathbf{x}_+)\;,$
			$L_+ := L$
5		**else**	**set** $L_+ := M\;,$
			$U_+ := U\;,$
			$\mathbf{x}_+ := \mathbf{x}$
	endif	**return**	

Advantages and disadvantages of the NB method

Advantages:

- Stopping rule
- Fast convergence. Average bracket reduction $\approx \frac{1}{2}$
- "Cold start" OK.
- Easy to program, no tricks.

Advantages and disadvantages of the NB method

Advantages:

- Stopping rule
- Fast convergence. Average bracket reduction $\approx \frac{1}{2}$
- "Cold start" OK.
- Easy to program, no tricks.

Disadvantages:

- Not valid (the rule for updating L) if f is ill-conditioned.

Quadratic functions

Theorem. Let

$$f(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}\cdot Q\mathbf{x} - \mathbf{c}\cdot\mathbf{x} + \gamma$$

Q positive definite with eigenvalues

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \ .$$

Then the NB method is valid for minimizing f if

$$\frac{\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} \geq \frac{1}{16}$$

or

$$\frac{\lambda_n}{\lambda_1} \leq \ \frac{1}{7-\sqrt{48}} = 13.9282$$

Why should we care?

• If Q is known, then the above f can be minimized explicitly. Howevere, all "nice" convex functions f are "almost quadratic" near their minima, with $Q = f''$, the Hessian matrix.

Why should we care?

• If Q is known, then the above f can be minimized explicitly. Howevere, all "nice" convex functions f are "almost quadratic" near their minima, with $Q = f''$, the Hessian matrix.

• If the quadratic function is the sum of squares of errors

$$f(\mathbf{x}) = \frac{1}{2}(A\mathbf{x} - \mathbf{b}) \cdot (A\mathbf{x} - \mathbf{b}) = \frac{1}{2}\mathbf{x} \cdot Q\mathbf{x} - \mathbf{c} \cdot \mathbf{x} + \gamma$$

with $Q = A^T A$, then the condition of Q,

$$\operatorname{cond} Q := \frac{\lambda_n(Q)}{\lambda_1 Q} = (\operatorname{cond} A)^2$$

i.e., if A is ill-conditioned, Q is much worse.

What can be done?

The quadratic

$$f(\mathbf{x}) = \frac{1}{2}\,\mathbf{x} \cdot Q\mathbf{x} - \mathbf{c} \cdot \mathbf{x} + \gamma$$

with positive definite Q, becomes a perfect sum of squares

$$f(\mathbf{y}) = \frac{1}{2}\,\mathbf{y} \cdot \mathbf{y} - \mathbf{c} \cdot Q^{-1/2}\mathbf{y} + \gamma$$

under the change-of-variables $\mathbf{y} := Q^{1/2}\mathbf{x}$.

For a convex function f near its minimum $\mathbf{x}^*$, the matrix Q is the (unknown) Hessian of f at the (unknown) $\mathbf{x}^*$. However, Q can be approximated locally during iterations, or the condition of f can be improved by dilation.

References

Y. Levin and A. B-I, Directional Newton Methods in n Variables, *Math. of Computation* **71**(2002), 237–250

Y. Levin and A. B-I, The Newton Bracketing Method for Convex Minimization, *Computational Optimization and Applications* **21**(2002), 213–229

Both can be downloaded at **http://benisrael.net/Newton.html**