# Lecture 8: Trust-region methods for unconstrained optimization

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

# Linesearch versus trust-region methods

$$\text{(UP):} \quad \text{minimize } f(x) \text{ subject to } x \in \mathbb{R}^n.$$

Linesearch methods: 'liberal' in the choice of search direction, keeping bad behaviour in control by choice of $\alpha^k$.

- choose descent direction $s^k$,
- compute stepsize $\alpha^k$ to reduce $f(x^k + \alpha s^k)$,
- update $x^{k+1} := x^k + \alpha^k s^k$.

# Linesearch versus trust-region methods

(UP): minimize $f(x)$ subject to $x \in \mathbb{R}^n$.

Linesearch methods: 'liberal' in the choice of search direction, keeping bad behaviour in control by choice of $\alpha^k$.

- choose descent direction $s^k$,
- compute stepsize $\alpha^k$ to reduce $f(x^k + \alpha s^k)$,
- update $x^{k+1} := x^k + \alpha^k s^k$.

Trust region (TR) methods: 'conservative' in the choice of search direction, so that a full stepsize along it may really reduce the objective.

- pick direction $s^k$ to reduce a "local model" of $f(x^k + s^k)$,
- accept $x^{k+1} := x^k + s^k$ if decrease in the model is also achieved by $f(x^k + s^k)$,
- else set $x^{k+1} := x^k$ and "refine" the model.

# Trust-region models for unconstrained problems

Approximate $f(x^k + s)$ by:

- linear model $\quad l_k(s) := f(x^k) + s^\top \nabla f(x^k) \qquad$ or
- quadratic model

$$q_k(s) := f(x^k) + s^\top \nabla f(x^k) + \frac{1}{2} s^\top \nabla^2 f(x^k) s.$$

Impediments:

models may not resemble $f(x^k + s)$ when $s$ is large,

models may be unbounded from below,

$*$ $l_k(s)$ always unbounded below (unless $\nabla f(x^k) = 0$)

$*$ $q_k(s)$ is always unbounded below if $\nabla^2 f(x^k)$ is negative definite or indefinite, and sometimes if $\nabla^2 f(x^k)$ is positive semidefinite.

# Trust region models and subproblem

Prevent bad approximations by trusting the model only in a trust region, defined by the trust region constraint

$$\|s\| \le \Delta_k, \tag{R}$$

for some "appropriate" radius $\Delta_k > 0$.

The constraint (R) also prevents $l_k$, $q_k$ from unboundedness!

# Trust region models and subproblem

Prevent bad approximations by trusting the model only in a trust region, defined by the trust region constraint

$$\|s\| \leq \Delta_k, \tag{R}$$

for some "appropriate" radius $\Delta_k > 0$.

The constraint (R) also prevents $l_k$, $q_k$ from unboundedness!

$\Longrightarrow$ the trust region subproblem

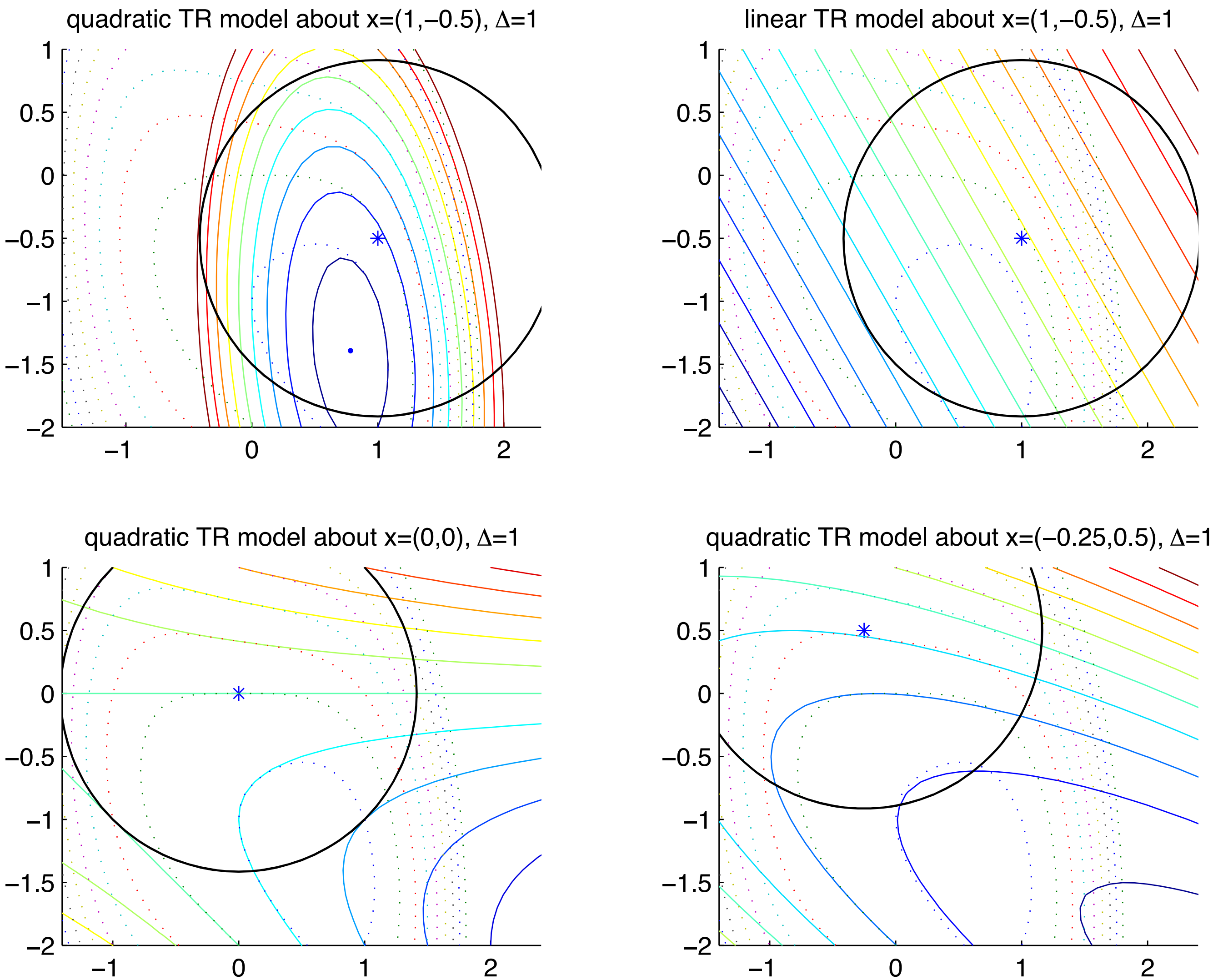
$$\min_{s \in \mathbb{R}^n} m_k(s) \quad \text{subject to} \quad \|s\| \leq \Delta_k, \tag{TR}$$

where $m_k := l_k,\ k \geq 0$, or $m_k := q_k,\ k \geq 0$.

- From now on, $m_k := q_k$.

(TR) easier to solve than (P). May even solve (TR) only approximately.

# Trust region models and subproblem - an example




Trust-region models of $f(x) = x_1^4 + x_1x_2 + (1 + x_2)^2$.

# Generic trust-region method

Let $s^k$ be a(n approximate) solution of (TR). Then

- predicted model decrease:
  $$m_k(0) - m_k(s^k) = f(x^k) - m_k(s^k).$$
- actual function decrease: $f(x^k) - f(x^k + s^k)$.

The trust region radius $\Delta_k$ is chosen based on the value of

$$\rho_k := \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - m_k(s^k)}.$$

If $\rho_k$ is not too smaller than $1$, $x^{k+1} := x^k + s^k$, $\Delta_{k+1} \geq \Delta_k$.

If $\rho_k$ close to or $\geq 1$, $\Delta_k$ is increased.

If $\rho_k \ll 1$, $x^{k+1} = x^k$ and $\Delta_k$ is reduced.

# A Generic Trust Region (GTR) method

Given $\Delta_0 > 0$, $x^0 \in \mathbb{R}^n$, $\epsilon > 0$. While $\|\nabla f(x^k)\| \geq \epsilon$, do:

**1.** Form the local quadratic model $m_k(s)$ of $f(x^k + s)$.

□

# A Generic Trust Region (GTR) method

Given $\Delta_0 > 0$, $x^0 \in \mathbb{R}^n$, $\epsilon > 0$. While $\|\nabla f(x^k)\| \geq \epsilon$, do:

**1.** Form the local quadratic model $m_k(s)$ of $f(x^k + s)$.

**2.** Solve (approximately) the (TR) subproblem for $s^k$ with $m_k(s^k) < f(x^k)$ ("sufficiently").

Compute $\rho_k := [f(x^k) - f(x^k + s^k)]/[f(x^k) - m_k(s^k)]$.

$\square$

# A Generic Trust Region (GTR) method

Given $\Delta_0 > 0$, $x^0 \in \mathbb{R}^n$, $\epsilon > 0$. While $\|\nabla f(x^k)\| \geq \epsilon$, do:

**1.** Form the local quadratic model $m_k(s)$ of $f(x^k + s)$.

**2.** Solve (approximately) the (TR) subproblem for $s^k$ with $m_k(s^k) < f(x^k)$ ("sufficiently").

Compute $\rho_k := [f(x^k) - f(x^k + s^k)]/[f(x^k) - m_k(s^k)]$.

**3.** If $\rho_k \geq 0.9$, then [very successful step]
set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := 2\Delta_k$.

$\square$

# A Generic Trust Region (GTR) method

Given $\Delta_0 > 0$, $x^0 \in \mathbb{R}^n$, $\epsilon > 0$. While $\|\nabla f(x^k)\| \geq \epsilon$, do:

**1.** Form the local quadratic model $m_k(s)$ of $f(x^k + s)$.

**2.** Solve (approximately) the (TR) subproblem for $s^k$ with $m_k(s^k) < f(x^k)$ ("sufficiently").

Compute $\rho_k := [f(x^k) - f(x^k + s^k)]/[f(x^k) - m_k(s^k)]$.

**3.** If $\rho_k \geq 0.9$, then [very successful step]

set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := 2\Delta_k$.

Else if $\rho_k \geq 0.1$, then [successful step]

set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := \Delta_k$.

$\square$

# A Generic Trust Region (GTR) method

Given $\Delta_0 > 0$, $x^0 \in \mathbb{R}^n$, $\epsilon > 0$. While $\|\nabla f(x^k)\| \geq \epsilon$, do:

**1.** Form the local quadratic model $m_k(s)$ of $f(x^k + s)$.

**2.** Solve (approximately) the (TR) subproblem for $s^k$ with $m_k(s^k) < f(x^k)$ ("sufficiently").

Compute $\rho_k := [f(x^k) - f(x^k + s^k)]/[f(x^k) - m_k(s^k)]$.

**3.** If $\rho_k \geq 0.9$, then [very successful step]

set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := 2\Delta_k$.

Else if $\rho_k \geq 0.1$, then [successful step]

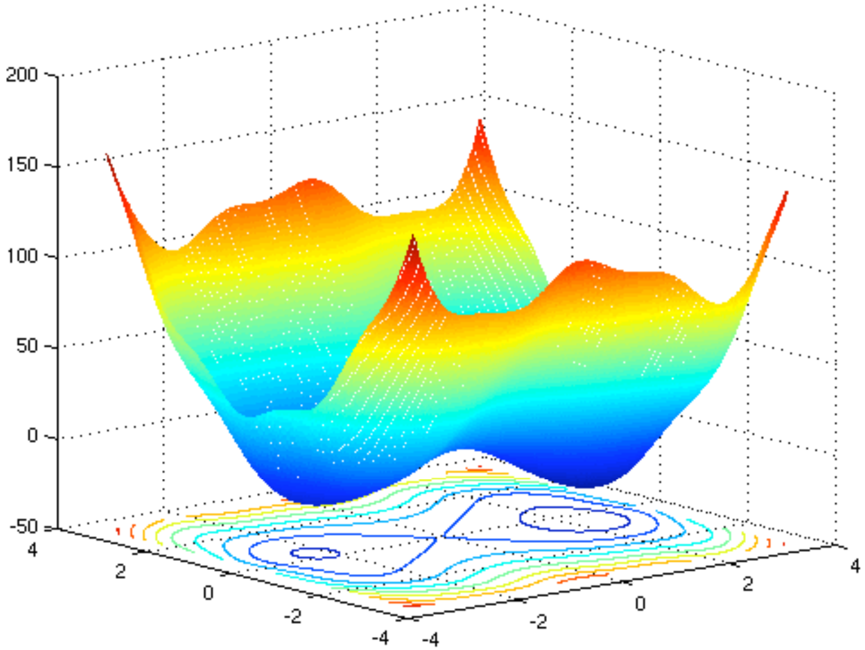
set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := \Delta_k$.

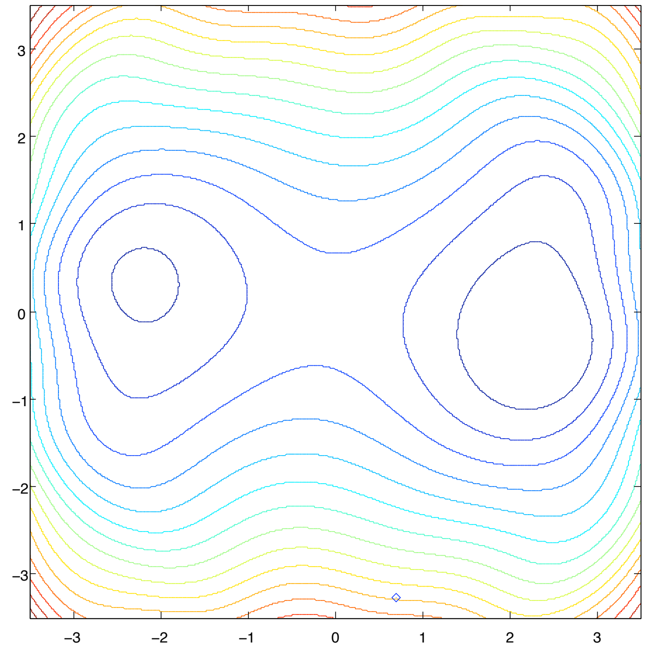
Else [unsuccessful step]

set $x^{k+1} = x^k$ and $\Delta_{k+1} := \frac{1}{2}\Delta_k$.

**4.** Let $k := k + 1$. □

$$\text{minimize}: \quad f(\alpha,\beta) = -10\alpha^2 + 10\beta^2 + 4\sin(\alpha\beta) - 2\alpha + \alpha^4$$

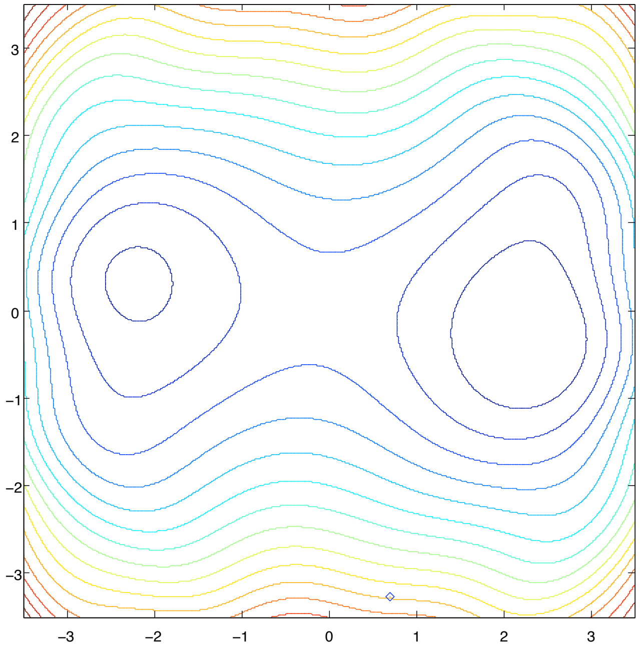


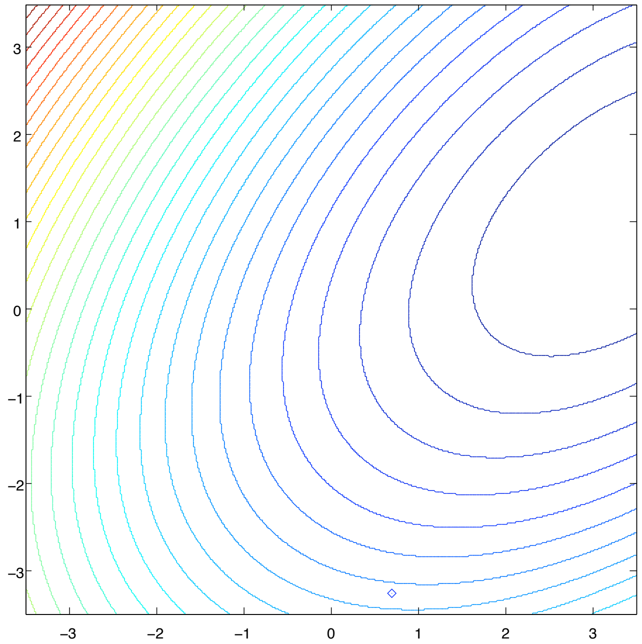

Two local minima: $(-2.20, 0.32)$ and $(2.30, -0.34)$

$$x_0 = (0.71, -3.27) \quad \text{and} \quad f(x_0) = 97.630$$

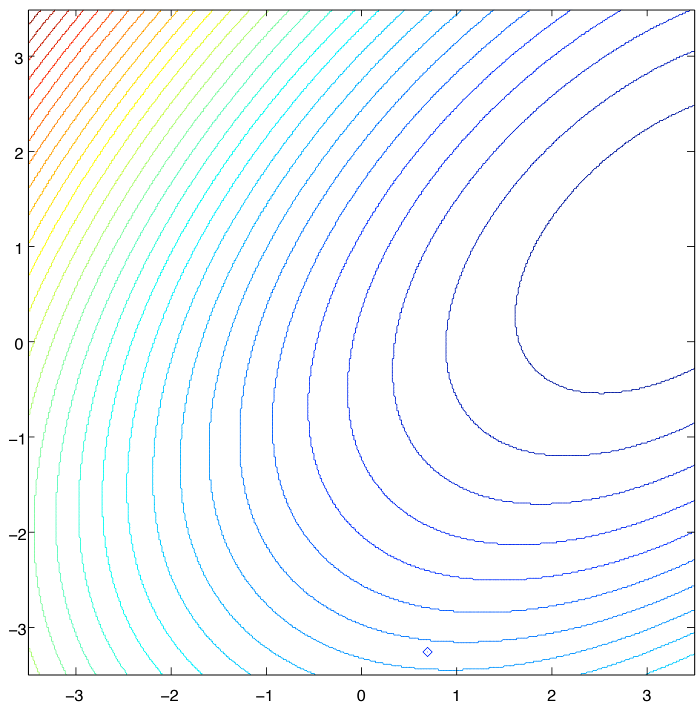
Contours of $f$

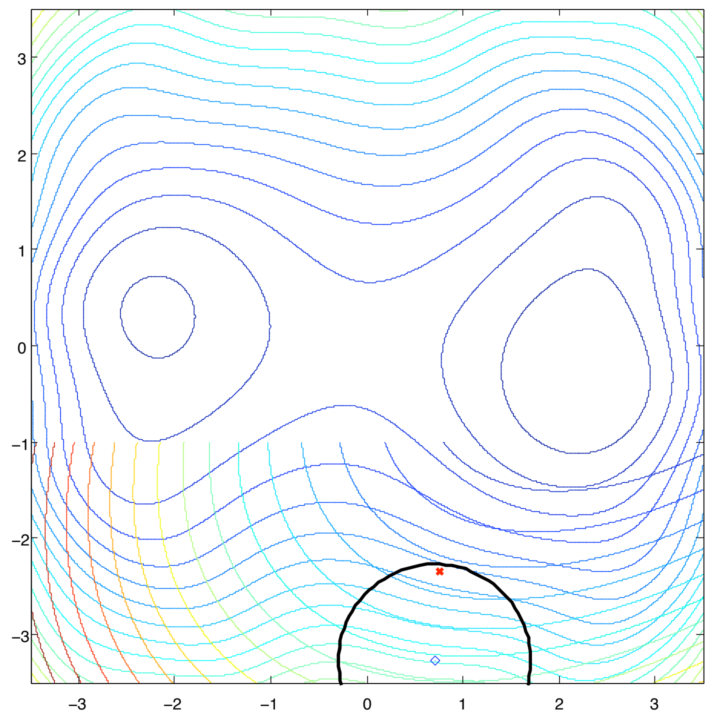



Contours of $m_0$ around $x_0$
(quadratic model)

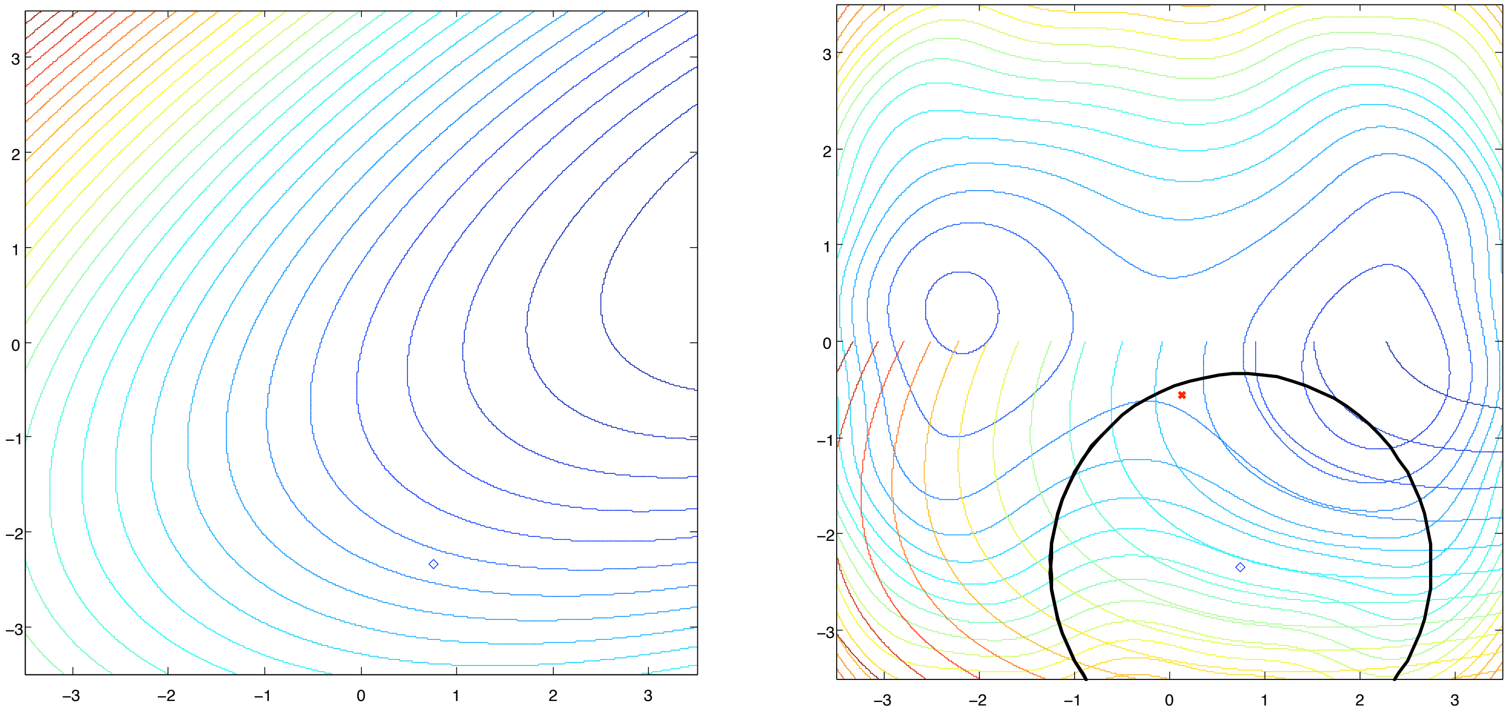


| $k$ | $\Delta_k$ | $s_k$ | $f(x_k + s_k)$ | $\Delta f / \Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | 1 | $(0.05, 0.93)$ | 43.742 | 0.998 | $x_0 + s_0$ |

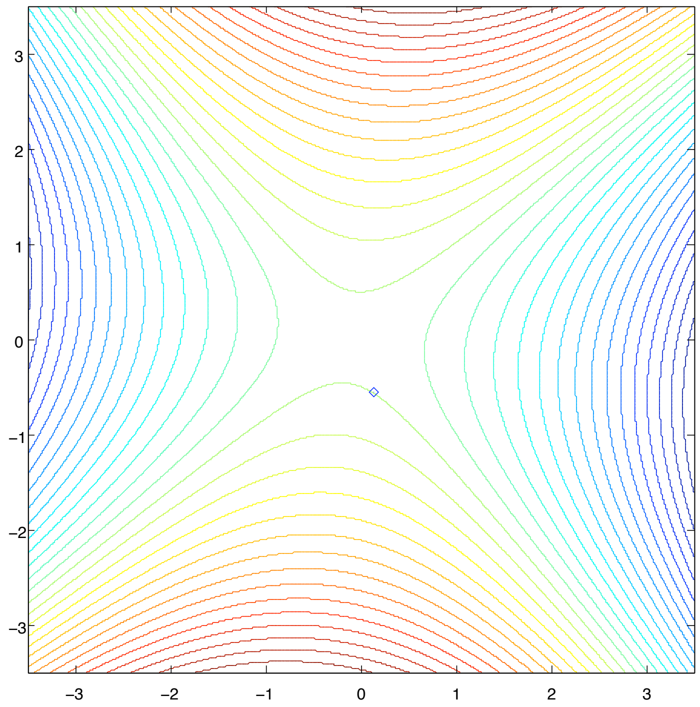


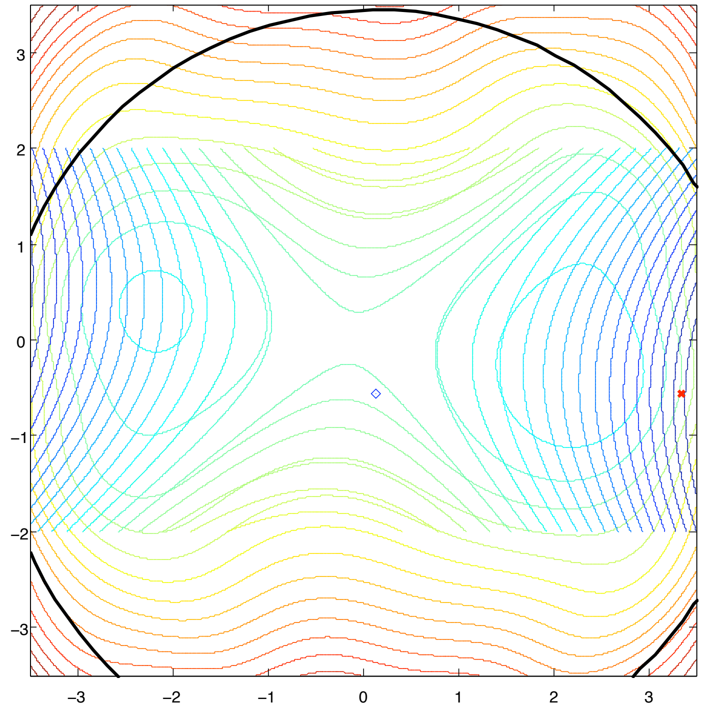



| $k$ | $\Delta_k$ | $s_k$ | $f(x_k + s_k)$ | $\Delta f / \Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | **1** | $(0.05, 0.93)$ | 43.742 | 0.998 | $x_0 + s_0$ |
| 1 | **2** | $(-0.62, 1.78)$ | 2.306 | 1.354 | $x_1 + s_1$ |

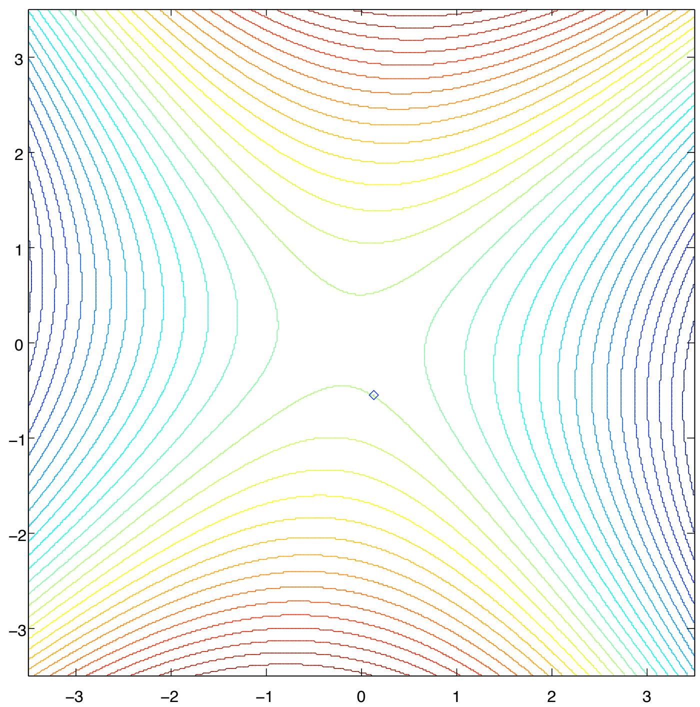


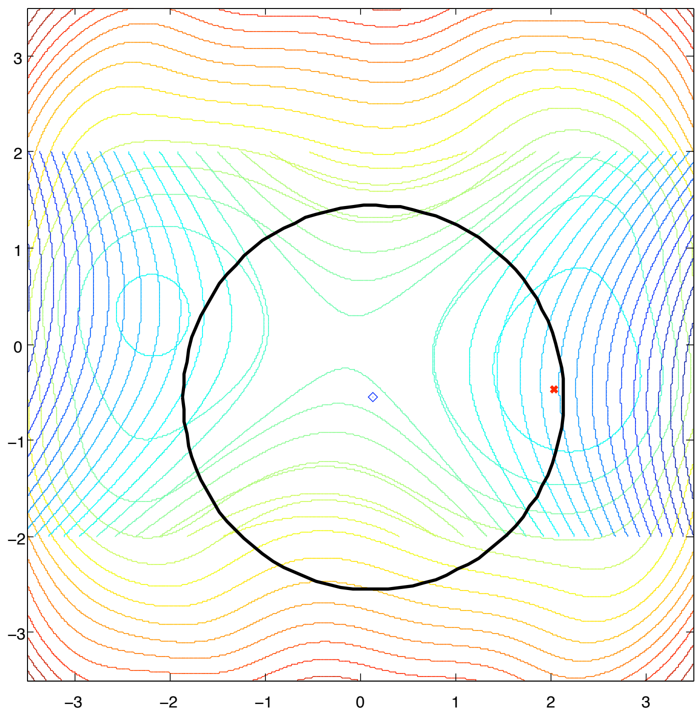
| $k$ | $\Delta_k$ | $s_k$ | $f(x_k + s_k)$ | $\Delta f / \Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | **1** | $(0.05, 0.93)$ | $43.742$ | $0.998$ | $x_0 + s_0$ |
| 1 | **2** | $(-0.62, 1.78)$ | $2.306$ | $1.354$ | $x_1 + s_1$ |
| 2 | **4** | $(3.21, 0.00)$ | $6.295$ | $-0.004$ | $x_2$ |



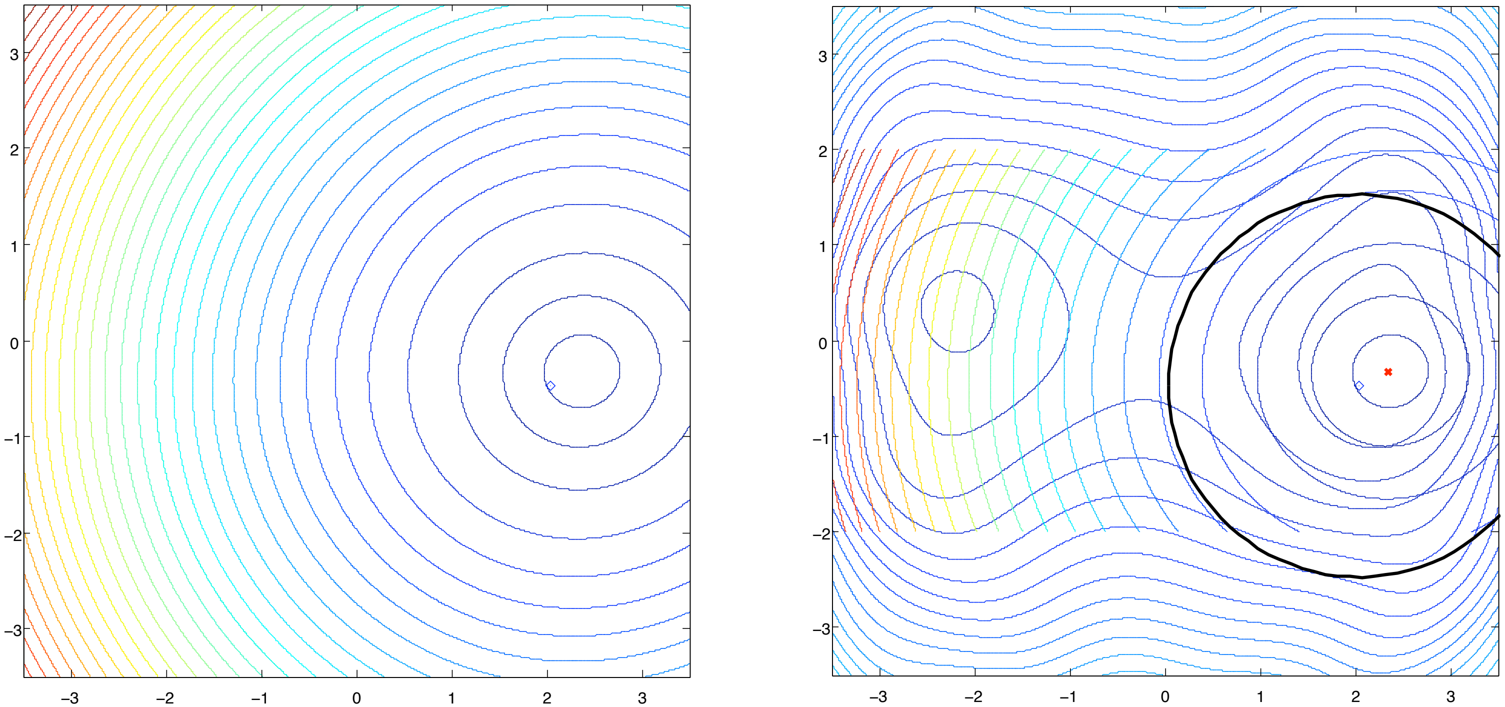



| $k$ | $\Delta_k$ | $s_k$ | $f(x_k + s_k)$ | $\Delta f / \Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | **1** | $(0.05, 0.93)$ | $43.742$ | $0.998$ | $x_0 + s_0$ |
| 1 | **2** | $(-0.62, 1.78)$ | $2.306$ | $1.354$ | $x_1 + s_1$ |
| 2 | **4** | $(3.21, 0.00)$ | $6.295$ | $-0.004$ | $x_2$ |
| 3 | **2** | $(1.90, 0.08)$ | $-29.392$ | $0.649$ | $x_2 + s_2$ |

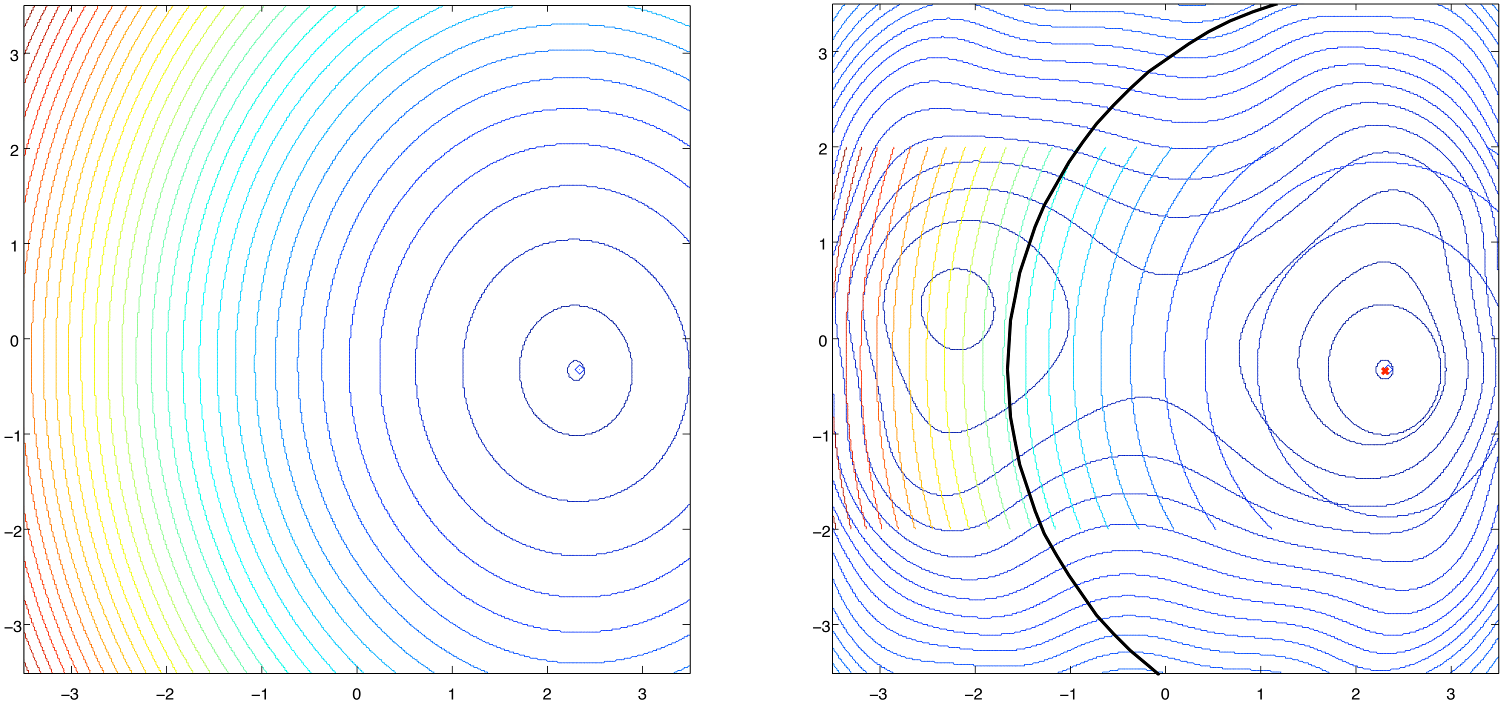


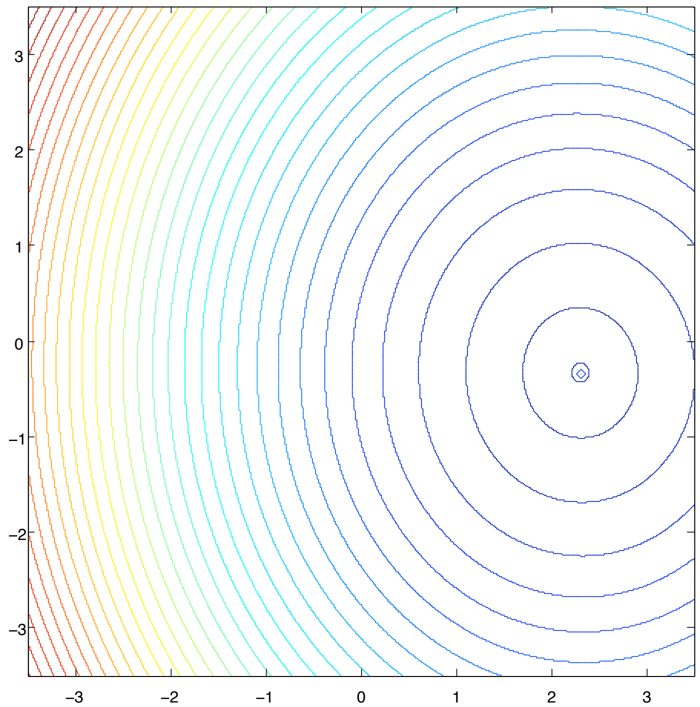
| $k$ | $\Delta_k$ | $s_k$ | $f(x_k + s_k)$ | $\Delta f / \Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | **1** | $(0.05, 0.93)$ | $43.742$ | $0.998$ | $x_0 + s_0$ |
| 1 | **2** | $(-0.62, 1.78)$ | $2.306$ | $1.354$ | $x_1 + s_1$ |
| 2 | **4** | $(3.21, 0.00)$ | $6.295$ | $-0.004$ | $x_2$ |
| 3 | **2** | $(1.90, 0.08)$ | $-29.392$ | $0.649$ | $x_2 + s_2$ |
| 4 | **2** | $(0.32, 0.15)$ | $-31.131$ | $0.857$ | $x_3 + s_3$ |

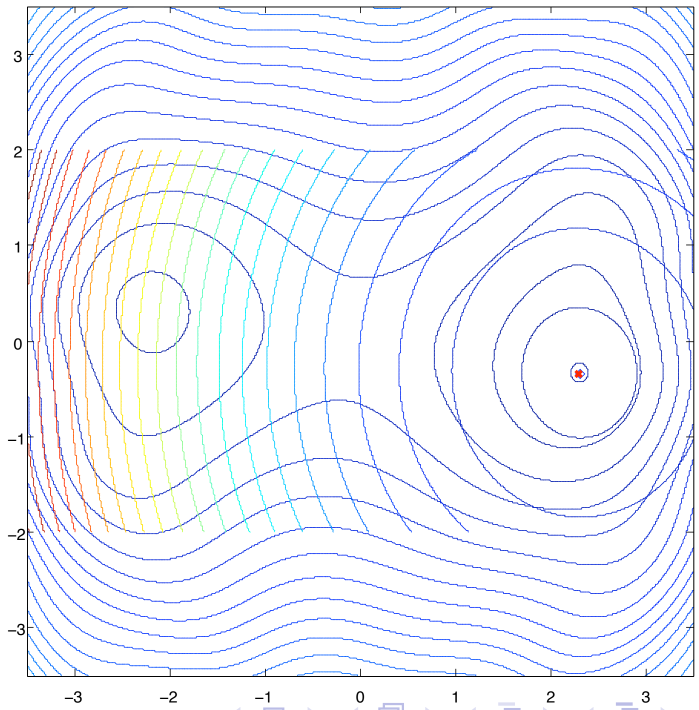


| $k$ | $\Delta_k$ | $s_k$ | $f(x_k+s_k)$ | $\Delta f/\Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | **1** | $(0.05, 0.93)$ | $43.742$ | $0.998$ | $x_0 + s_0$ |
| 1 | **2** | $(-0.62, 1.78)$ | $2.306$ | $1.354$ | $x_1 + s_1$ |
| 2 | **4** | $(3.21, 0.00)$ | $6.295$ | $-0.004$ | $x_2$ |
| 3 | **2** | $(1.90, 0.08)$ | $-29.392$ | $0.649$ | $x_2 + s_2$ |
| 4 | **2** | $(0.32, 0.15)$ | $-31.131$ | $0.857$ | $x_3 + s_3$ |
| 5 | **4** | $(-0.03, -0.02)$ | $-31.176$ | $1.009$ | $x_4 + s_4$ |

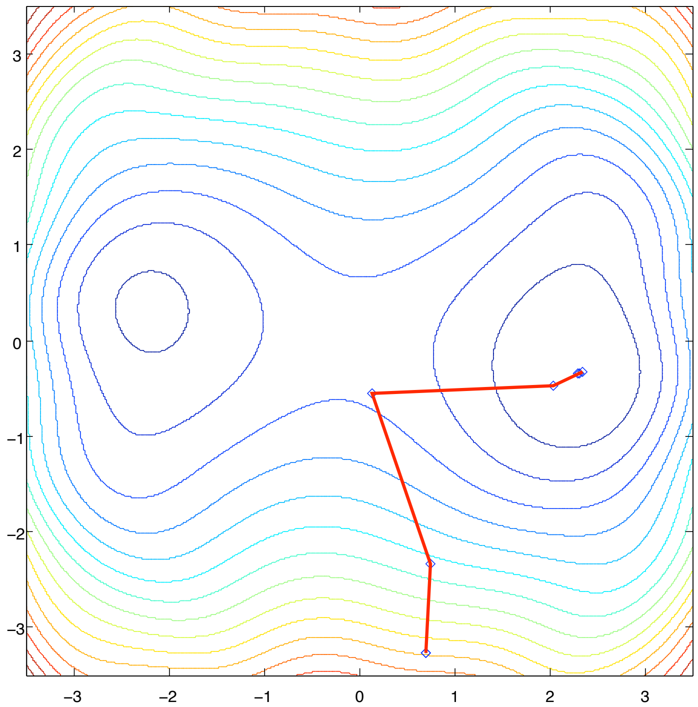


| $k$ | $\Delta_k$ | $s_k$ | $f(x_k + s_k)$ | $\Delta f / \Delta m_k$ | $x_{k+1}$ |
|---|---|---|---|---|---|
| 0 | **1** | $(0.05, 0.93)$ | 43.742 | 0.998 | $x_0 + s_0$ |
| 1 | **2** | $(-0.62, 1.78)$ | 2.306 | 1.354 | $x_1 + s_1$ |
| 2 | **4** | $(3.21, 0.00)$ | 6.295 | $-0.004$ | $x_2$ |
| 3 | **2** | $(1.90, 0.08)$ | $-29.392$ | 0.649 | $x_2 + s_2$ |
| 4 | **2** | $(0.32, 0.15)$ | $-31.131$ | 0.857 | $x_3 + s_3$ |
| 5 | **4** | $(-0.03, -0.02)$ | $-31.176$ | 1.009 | $x_4 + s_4$ |
| 6 | **8** | $(-0.02, 0.00)$ | $-31.179$ | 1.013 | $x_5 + s_5$ |

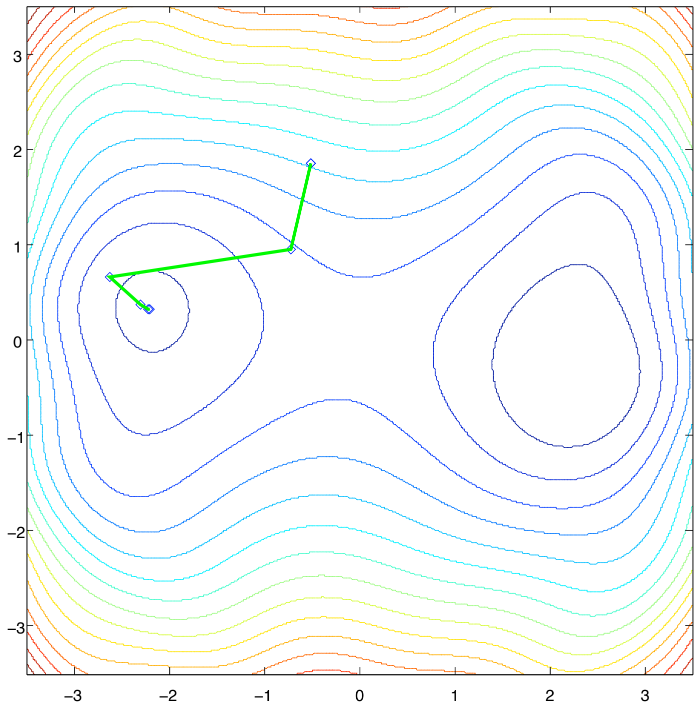


Path of iterates:




From another $x_0$:



# Trust-region methods

- Other sensible values of the parameters of the GTR are possible.

"Solving" the (TR) subproblem

$$\min_{s \in \mathbb{R}^n} m_k(s) \quad \text{subject to} \quad \|s\| \le \Delta_k, \tag{TR}$$

... exactly or even approximately may imply work.

Want "minimal" condition of "sufficient decrease" in the model that ensures global convergence of the TR method (the Cauchy cond.). In practice, we (usually) do much better than this condition!

Example of applying a trust-region method: [Sartenaer, 2008].

- approximate solution of (TR) subproblem: better than Cauchy, but not exact.

- notation: $\Delta f / \Delta m_k \equiv \rho_k$.

# The Cauchy point of the (TR) subproblem

• recall the steepest descent method has strong (theoretical) global convergence properties; same will hold for TR method with SD direction.

"minimal" condition of "sufficient decrease" in the model: require

$$m_k(s^k) \leq m_k(s_c^k) \text{ and } \|s^k\| \leq \Delta_k,$$

where $s_c^k := -\alpha_c^k \nabla f(x^k)$, with

$$\alpha_c^k := \text{arg} \min_{\alpha>0} m_k(-\alpha \nabla f(x^k)) \text{ subject to } \|\alpha \nabla f(x^k)\| \leq \Delta_k.$$

[i.e. a linesearch along steepest descent direction is applied to $m_k$ at $x^k$ and is restricted to the trust region.] Easy:

$$\alpha_c^k := \text{arg} \min_{\alpha} m_k(-\alpha \nabla f(x^k)) \text{ subject to } 0 < \alpha \leq \frac{\Delta_k}{\|\nabla f(x^k)\|}.$$

• $y_c^k := x^k + s_c^k$ is the Cauchy point.

# Global convergence of the GTR method

Theorem 11 (GTR global convergence)

Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and bounded below on $\mathbb{R}^n$. Let $\nabla f$ be Lipschitz continuous on $\mathbb{R}^n$. Let $\{x^k\}$ be generated by the generic trust region (GTR) method, and let the computation of $s^k$ be such that $m_k(s^k) \le m_k(s_c^k)$ for all $k$. Then either

$$\text{there exists } k \ge 0 \text{ such that } \nabla f(x^k) = 0$$

or

$$\lim_{k\to\infty} \|\nabla f(x^k)\| = 0.$$

We (only) sketch the proof of $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$ (which also implies finite termination of GTR) next.