Weighted Trust-Region Methods

Johannes J. Brust[1] Philip E. Gill[1]

[1] Department of Mathematics
University of California, San Diego, CA

US & Mexico Workshop on Optimization and its Applications

January 9 -13th, 2023

Outline

Problem formulation

Method

Numerical Experiments

Conclusions

Problem Formulation

Nonlinear unconstrained optimization

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\mathbf{x})$$

where $f : \mathbb{R}^n \to \mathbb{R}$.

Problem Formulation

Nonlinear unconstrained optimization

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\mathbf{x})$$

where $f : \mathbb{R}^n \to \mathbb{R}$.

Assumptions:

- f is twice continuously differentiable
- Gradients $\nabla f(\mathbf{x})$ are available
- Second derivatives are unavailable

Trust-Region Step

Iterates are updated in a trust-region method: $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$

Trust-Region Step

Iterates are updated in a trust-region method: $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$

A quadratic subproblem defines a step:

$$\underset{\|\mathbf{s}\| \leq \Delta_k}{\operatorname{argmin}} \; \mathbf{s}^\top \mathbf{g}_k + \frac{1}{2} \mathbf{s}^\top B_k \mathbf{s}$$

$$0 < \Delta_k \text{ (radius)}, \quad \mathbf{g}_k = \nabla f(\mathbf{x}_k), \quad B_k \text{ (symmetric } n \times n)$$

Trust-Region Step

Iterates are updated in a trust-region method: $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$

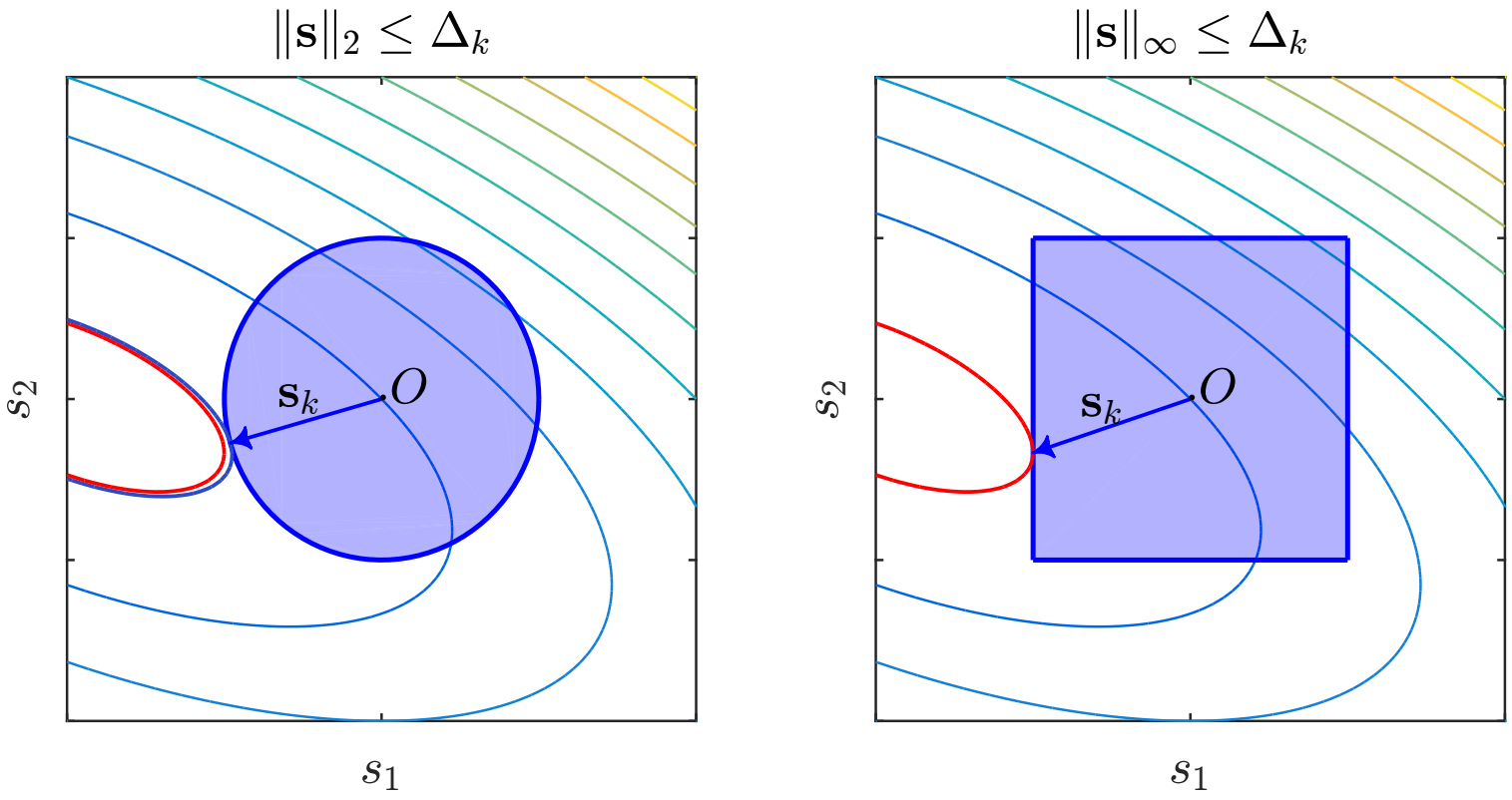
A quadratic subproblem defines a step:

$$\underset{\|\mathbf{s}\| \leq \Delta_k}{\operatorname{argmin}} \; \mathbf{s}^\top \mathbf{g}_k + \frac{1}{2} \mathbf{s}^\top B_k \mathbf{s}$$

$$0 < \Delta_k \text{ (radius)}, \quad \mathbf{g}_k = \nabla f(\mathbf{x}_k), \quad B_k \text{ (symmetric } n \times n)$$

Typical norms are the two-norm or infinity-norm.

Typical subproblem norms



Computing the trust-region step is normally challenging.

Related Work

For efficiency, often approximate solutions to the trust-region subproblem are effective, with a family of methods:

[Moré and Sorenson, ’81]: Sequence of Cholesky factorizations

[Steihaug, ’83]: Truncated conjuage-gradient

[Gertz, ’04]: Infinity norm trust-region quasi-Newton

[Nocedal and Wright, ’06]: Dogleg method

Related Work

For efficiency, often approximate solutions to the trust-region subproblem are effective, with a family of methods:

[Moré and Sorenson, ’81]: Sequence of Cholesky factorizations

[Steihaug, ’83]: Truncated conjuage-gradient

[Gertz, ’04]: Infinity norm trust-region quasi-Newton

[Nocedal and Wright, ’06]: Dogleg method

Even these methods can be computationally intensive for large problems, or not applicable to indefinite subproblems.

Method

Suppose a stable symmetric indefinite factorization is obtained

$$B_k = L_k D_k L_k^\top,$$

L_k is lower triangular with **normalized columns**, D_k is diagonal.

Method

Suppose a stable symmetric indefinite factorization is obtained

$$B_k = L_k D_k L_k^\top,$$

L_k is lower triangular with **normalized columns**, D_k is diagonal.

Properties:

- D_k and B_k share the same inertia
- Solves with the factorization are efficient
- Effective rank-1 updates to the factorization

Method

Use the factorization for a change of variable: $\mathbf{v} = L_k^\top \mathbf{s}$ and

$$\mathbf{s}^\top B_k \mathbf{s} = \mathbf{s}^\top L_k D_k L_k^\top \mathbf{s} = \mathbf{v}^\top D_k \mathbf{v}$$

A diagonal Hessian in the new variables.

Method

Use the factorization for a change of variable: $\mathbf{v} = L_k^\top \mathbf{s}$ and

$$\mathbf{s}^\top B_k \mathbf{s} = \mathbf{s}^\top L_k D_k L_k^\top \mathbf{s} = \mathbf{v}^\top D_k \mathbf{v}$$

A diagonal Hessian in the new variables.

Consider the weighted norm

$$\|\mathbf{v}\| = \|L_k^\top \mathbf{s}\| \leq \Delta_k$$

Method

Use the factorization for a change of variable: $\mathbf{v} = L_k^\top \mathbf{s}$ and

$$\mathbf{s}^\top B_k \mathbf{s} = \mathbf{s}^\top L_k D_k L_k^\top \mathbf{s} = \mathbf{v}^\top D_k \mathbf{v}$$

A diagonal Hessian in the new variables.

Consider the weighted norm

$$\|\mathbf{v}\| = \|L_k^\top \mathbf{s}\| \leq \Delta_k$$

The trust-region subproblem simplifies this way.

Method

Use the factorization for a change of variable: $\mathbf{v} = L_k^\top \mathbf{s}$ and

$$\mathbf{s}^\top B_k \mathbf{s} = \mathbf{s}^\top L_k D_k L_k^\top \mathbf{s} = \mathbf{v}^\top D_k \mathbf{v}$$

A diagonal Hessian in the new variables.

Consider the weighted norm

$$\|\mathbf{v}\| = \|L_k^\top \mathbf{s}\| \leq \Delta_k$$

The trust-region subproblem simplifies this way.

Note: $\mathbf{s}^\top \mathbf{g}_k = \mathbf{v}^\top L_k^{-1} \mathbf{g}_k$

Method

The weighted trust-region subproblem (WTR)

$$\min_{\|L_k^\top \mathbf{s}\| \leq \Delta_k} \mathbf{s}^\top \mathbf{g}_k + \frac{1}{2}\mathbf{s}^\top B_k \mathbf{s} \;=\; \min_{\|\mathbf{v}\| \leq \Delta_k} \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2}\mathbf{v}^\top D_k \mathbf{v}$$

Method

The weighted trust-region subproblem (WTR)

$$\min_{\|L_k^\top \mathbf{s}\| \leq \Delta_k} \mathbf{s}^\top \mathbf{g}_k + \frac{1}{2}\mathbf{s}^\top B_k \mathbf{s} \;=\; \min_{\|\mathbf{v}\| \leq \Delta_k} \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2}\mathbf{v}^\top D_k \mathbf{v}$$

Properties:

- The solution $\mathbf{v}_k$ to (WTR) can be found straightforwardly
- The step from a triangular solve $\mathbf{s}_k = L_k^{-\top}\mathbf{v}_k$

Method

The weighted trust-region subproblem (WTR)

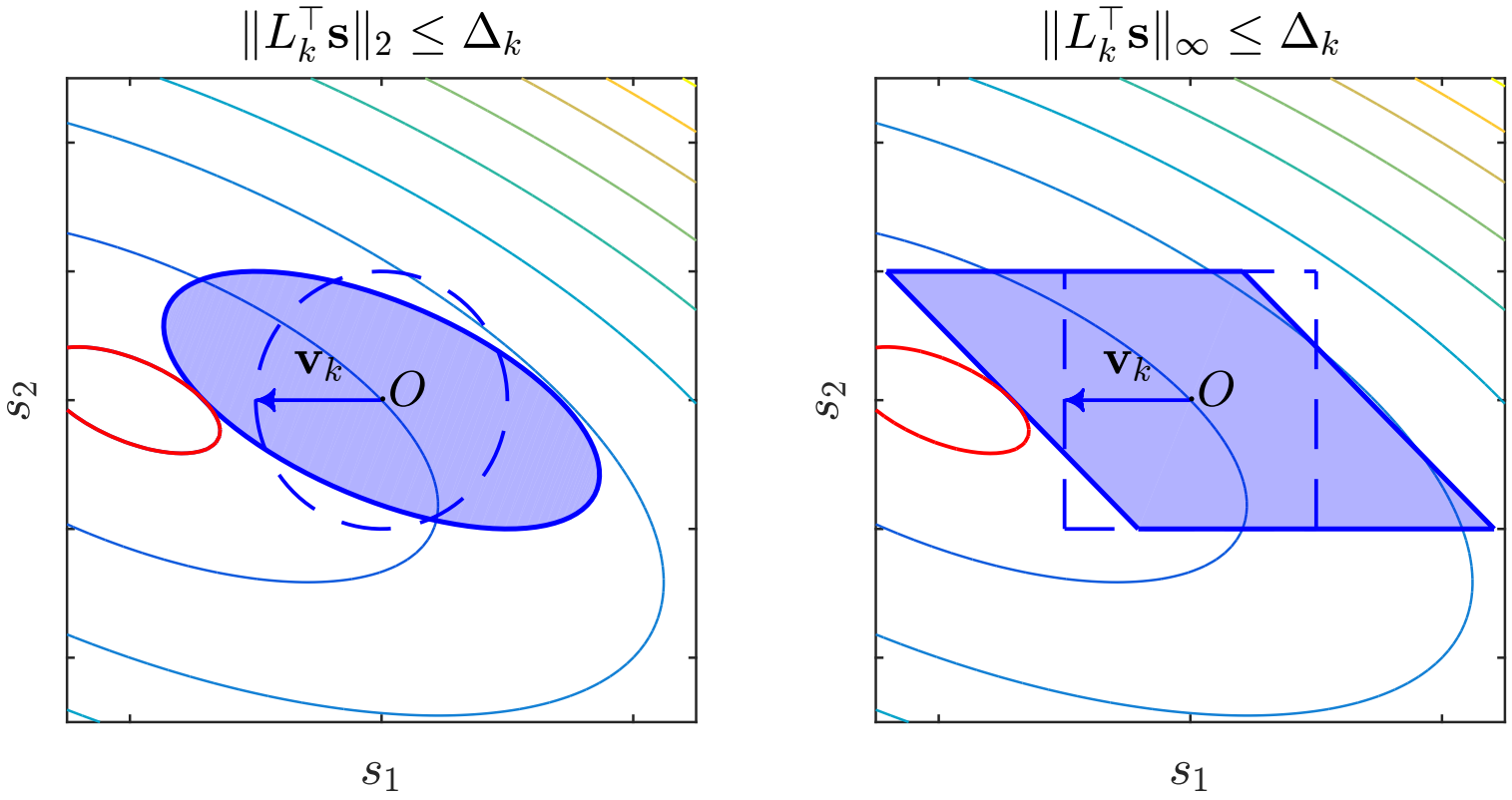
$$\min_{\|L_k^\top \mathbf{s}\| \leq \Delta_k} \mathbf{s}^\top \mathbf{g}_k + \frac{1}{2}\mathbf{s}^\top B_k \mathbf{s} \;=\; \min_{\|\mathbf{v}\| \leq \Delta_k} \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2}\mathbf{v}^\top D_k \mathbf{v}$$

Properties:

- The solution $\mathbf{v}_k$ to (WTR) can be found straightforwardly
- The step from a triangular solve $\mathbf{s}_k = L_k^{-\top}\mathbf{v}_k$

Different weighted norms are possible, e.g. $\|L_k^\top \mathbf{s}\|_2$ or $\|L_k^\top \mathbf{s}\|_\infty$

Weighted Subproblems: WTR



Computing the weighted trust-region step $L^\top \mathbf{s}_k = \mathbf{v}_k$ is less challenging.

Weighted L2 norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_2 \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Weighted L2 norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_2 \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Find a pair $(\mathbf{v}_k, \sigma_k)$ that satisfies the optimality conditions:
$\sigma_k \geq 0$,

$$(D_k + \sigma_k I) \succeq 0, \quad (D_k + \sigma_k I)\mathbf{v}_k = -L_k^{-1}\mathbf{g}_k, \quad \sigma_k(\|\mathbf{v}_k\|_2 - \Delta_k) = 0$$

Weighted L2 norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_2 \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Find a pair $(\mathbf{v}_k, \sigma_k)$ that satisfies the optimality conditions:
$\sigma_k \geq 0$,

$$(D_k + \sigma_k I) \succeq 0, \quad (D_k + \sigma_k I)\mathbf{v}_k = -L_k^{-1}\mathbf{g}_k, \quad \sigma_k(\|\mathbf{v}_k\|_2 - \Delta_k) = 0$$

A **1D Newton iteration** can efficiently determine σ_k and $\mathbf{v}_k$, since D_k is diagonal.

Weighted L2 norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_2 \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Find a pair $(\mathbf{v}_k, \sigma_k)$ that satisfies the optimality conditions:
$\sigma_k \geq 0$,

$$(D_k + \sigma_k I) \succeq 0, \quad (D_k + \sigma_k I)\mathbf{v}_k = -L_k^{-1}\mathbf{g}_k, \quad \sigma_k(\|\mathbf{v}_k\|_2 - \Delta_k) = 0$$

A **1D Newton iteration** can efficiently determine σ_k and $\mathbf{v}_k$, since D_k is diagonal.

Obtain the step from a triangular solve $\mathbf{s}_k = L_k^{-\top}\mathbf{v}_k$.

Weighted INF norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_\infty \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Weighted INF norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_\infty \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Note: Since D_k is diagonal the problem is separable

Weighted INF norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_\infty \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Note: Since D_k is diagonal the problem is separable

The analytic solution, when D_k is positive definite is

$$(\mathbf{v}_k)_i = \min(\Delta_k, \max(-\Delta_k, -(D_k^{-1} L_k^{-1} \mathbf{g}_k)_i))$$

Weighted INF norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_\infty \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Note: Since D_k is diagonal the problem is separable

The analytic solution, when D_k is positive definite is

$$(\mathbf{v}_k)_i = \min(\Delta_k, \max(-\Delta_k, -(D_k^{-1} L_k^{-1} \mathbf{g}_k)_i))$$

Obtain the step from a triangular solve $\mathbf{s}_k = L_k^{-\top} \mathbf{v}_k$.

Weighted INF norm

Solve the trust-region subproblem

$$\underset{\|\mathbf{v}\|_\infty \leq \Delta_k}{\text{minimize}} \;\; \mathbf{v}^\top L_k^{-1} \mathbf{g}_k + \frac{1}{2} \mathbf{v}^\top D_k \mathbf{v}.$$

Note: Since D_k is diagonal the problem is separable

The analytic solution, when D_k is positive definite is

$$(\mathbf{v}_k)_i = \min(\Delta_k, \max(-\Delta_k, -(D_k^{-1} L_k^{-1} \mathbf{g}_k)_i))$$

Obtain the step from a triangular solve $\mathbf{s}_k = L_k^{-\top} \mathbf{v}_k$.
(An analytic solution is also found when D_k is indefinite)

Theoretical Bounds

Bounds of the weighted norms: $\|L_k^\top \mathbf{s}\|_2$ and $\|L_k^\top \mathbf{s}\|_\infty$

Theoretical Bounds

Bounds of the weighted norms: $\|L_k^\top \mathbf{s}\|_2$ and $\|L_k^\top \mathbf{s}\|_\infty$

Lower triangular matrix with normalized columns

$$L_k^\top = \begin{bmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ & l_{22} & l_{32} & l_{42} \\ & & l_{33} & l_{43} \\ & & & 1 \end{bmatrix}, \qquad \mathrm{diag}(L_k^\top L_k) = I$$

Theoretical Bounds

Bounds of the weighted norms: $\|L_k^\top \mathbf{s}\|_2$ and $\|L_k^\top \mathbf{s}\|_\infty$

Lower triangular matrix with normalized columns

$$L_k^\top = \begin{bmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ & l_{22} & l_{32} & l_{42} \\ & & l_{33} & l_{43} \\ & & & 1 \end{bmatrix}, \qquad \mathrm{diag}(L_k^\top L_k) = I$$

For σ_1 the smallest singular value of L_k then

$$\sigma_1 \|\mathbf{s}\|_2 \leq \|L_k^\top \mathbf{s}\|_2 \leq \sqrt{n}\|\mathbf{s}\|_2,$$

$$|s_n| \leq \|L_k^\top \mathbf{s}\|_\infty \leq n\|\mathbf{s}\|_\infty$$

Implementation

The Hessian is estimated by a BFGS matrix.

$$B_{k+1} = B_k - \frac{1}{\mathbf{s}_k^\top B_k \mathbf{s}_k} B_k \mathbf{s}_k \mathbf{s}_k^\top B_k + \frac{1}{\mathbf{s}_k^\top \mathbf{y}_k} \mathbf{y}_k \mathbf{y}_k^\top$$

Here, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ and $\mathbf{s}_k^\top \mathbf{y}_k > 0$ ensures positive definiteness.

Implementation

The Hessian is estimated by a BFGS matrix.

$$B_{k+1} = B_k - \frac{1}{\mathbf{s}_k^\top B_k \mathbf{s}_k} B_k \mathbf{s}_k \mathbf{s}_k^\top B_k + \frac{1}{\mathbf{s}_k^\top \mathbf{y}_k} \mathbf{y}_k \mathbf{y}_k^\top$$

Here, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ and $\mathbf{s}_k^\top \mathbf{y}_k > 0$ ensures positive definiteness.

[Gill, Saunders et *al.*, '74]: Updates for the factorization

$$L_{k+1} D_{k+1} L_{k+1}^\top = L_k D_k L_k^\top - [\text{rank-1}] + [\text{rank-1}]$$

Implementation

The Hessian is estimated by a BFGS matrix.

$$B_{k+1} = B_k - \frac{1}{\mathbf{s}_k^\top B_k \mathbf{s}_k} B_k \mathbf{s}_k \mathbf{s}_k^\top B_k + \frac{1}{\mathbf{s}_k^\top \mathbf{y}_k} \mathbf{y}_k \mathbf{y}_k^\top$$

Here, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ and $\mathbf{s}_k^\top \mathbf{y}_k > 0$ ensures positive definiteness.

[Gill, Saunders et *al.*, '74]: Updates for the factorization

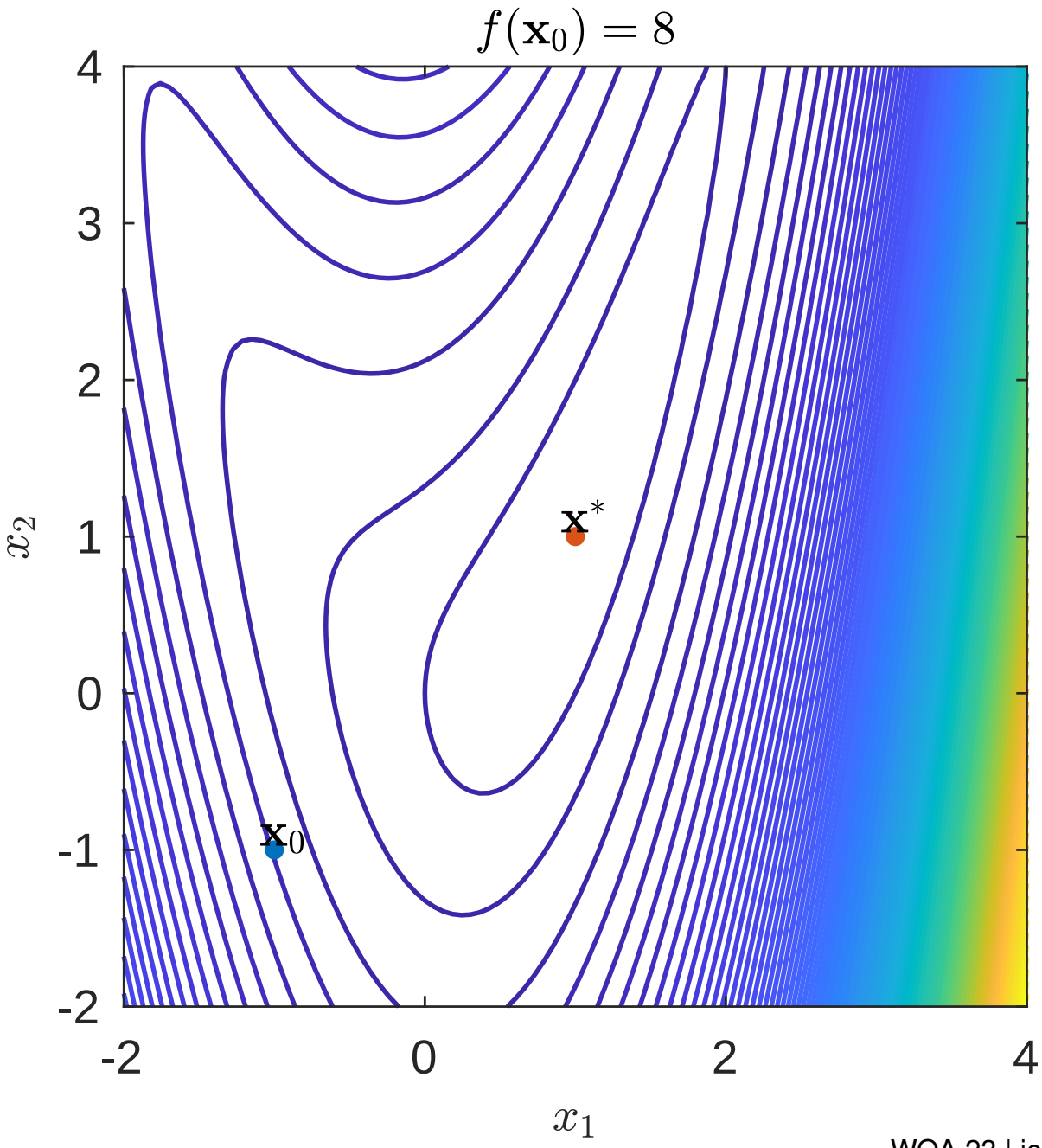
$$L_{k+1} D_{k+1} L_{k+1}^\top = L_k D_k L_k^\top - [\text{rank-1}] + [\text{rank-1}]$$

Other Hessian estimates (e.g., SR1) are possible.

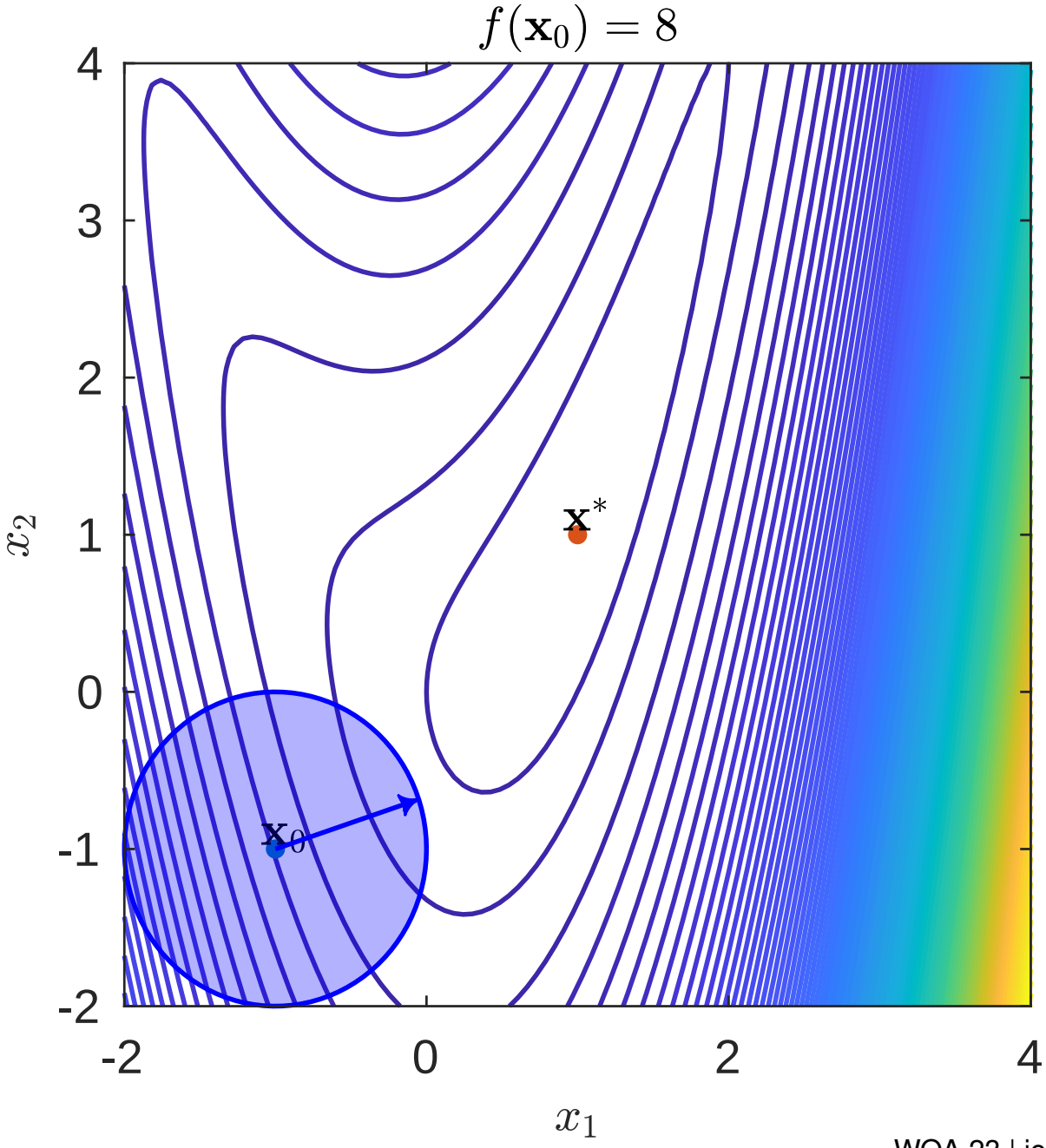
Example

Example: Rosenbrock 2D function

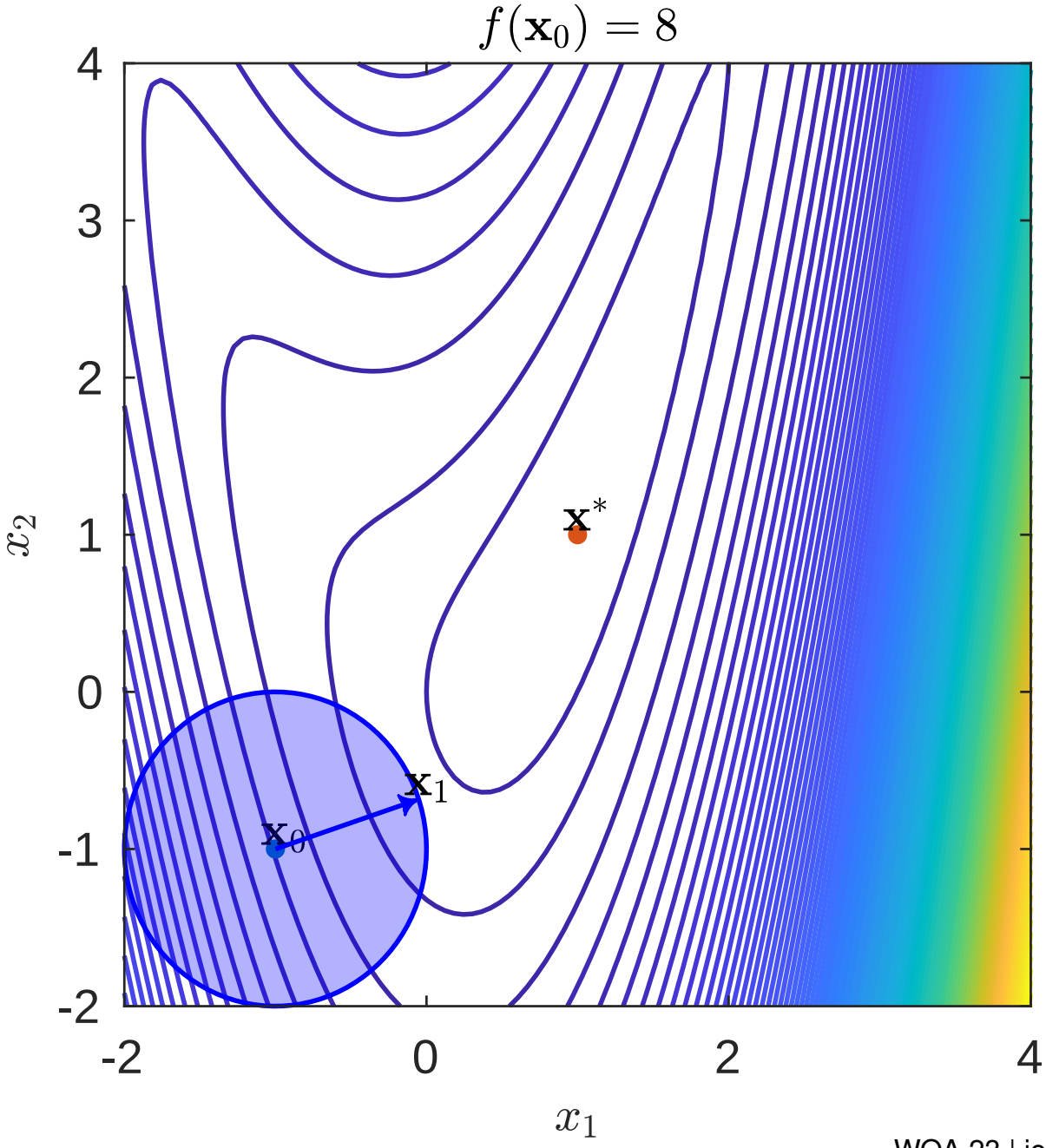
WTR-L2



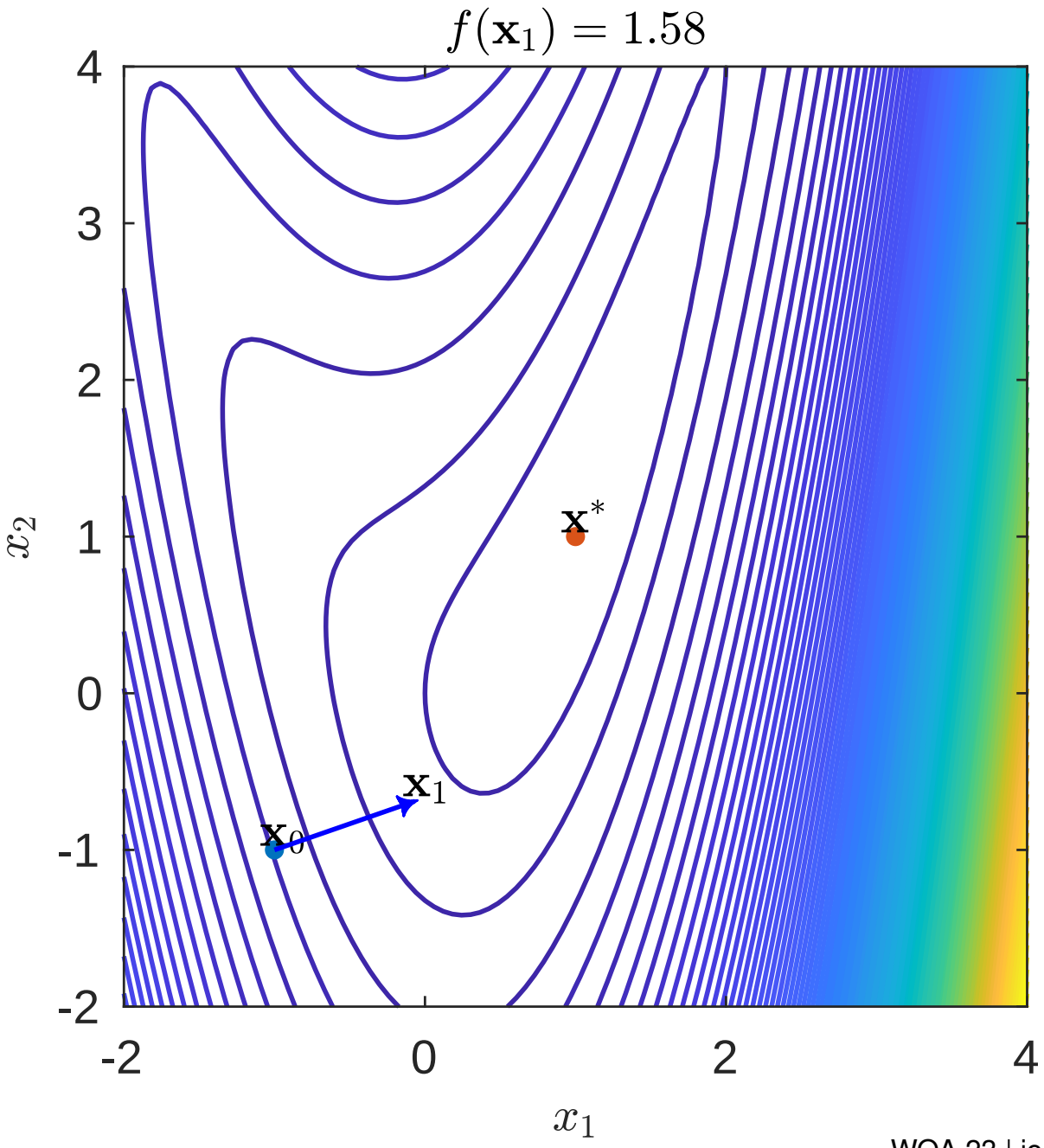
WTR-L2



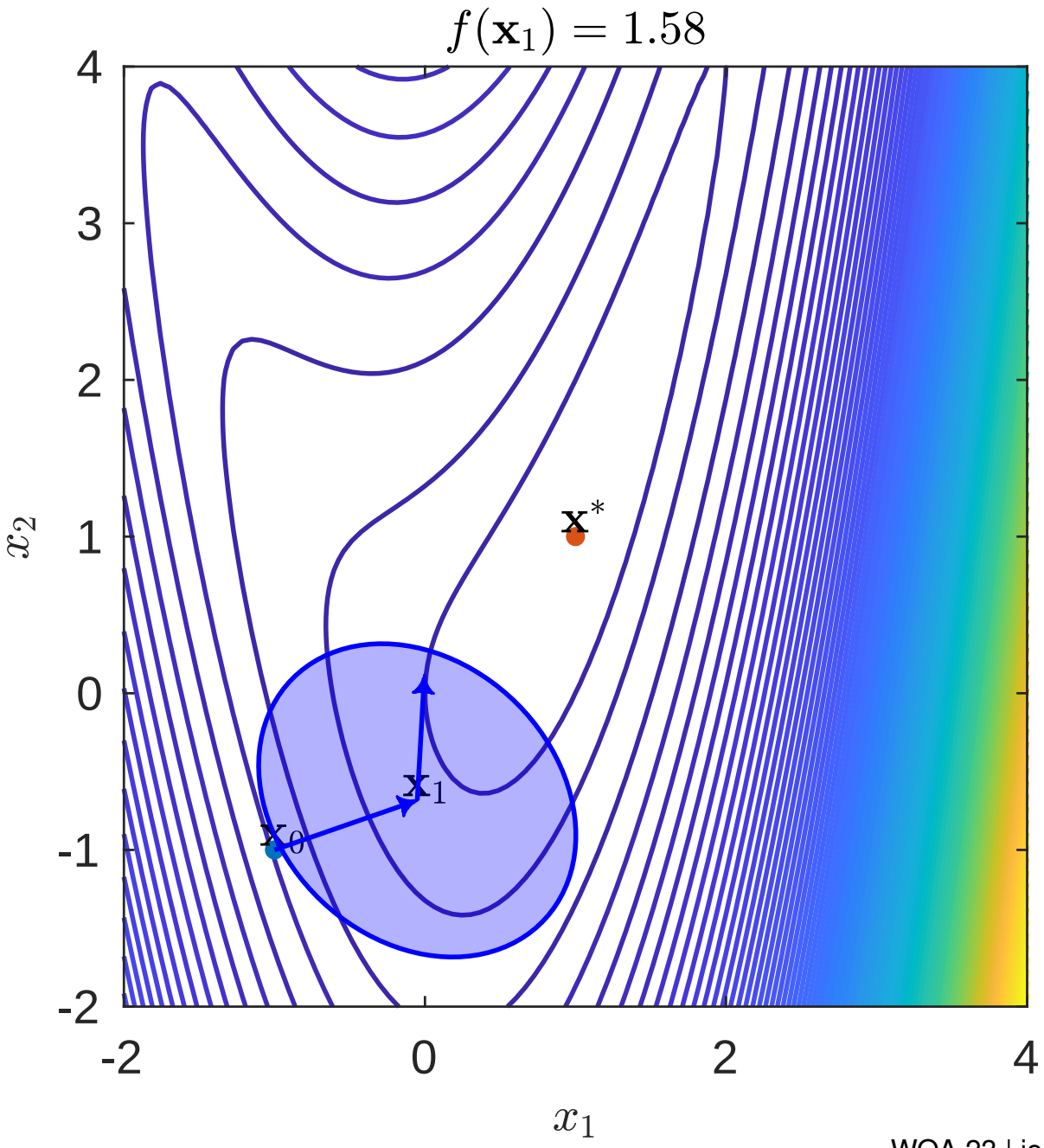
WTR-L2



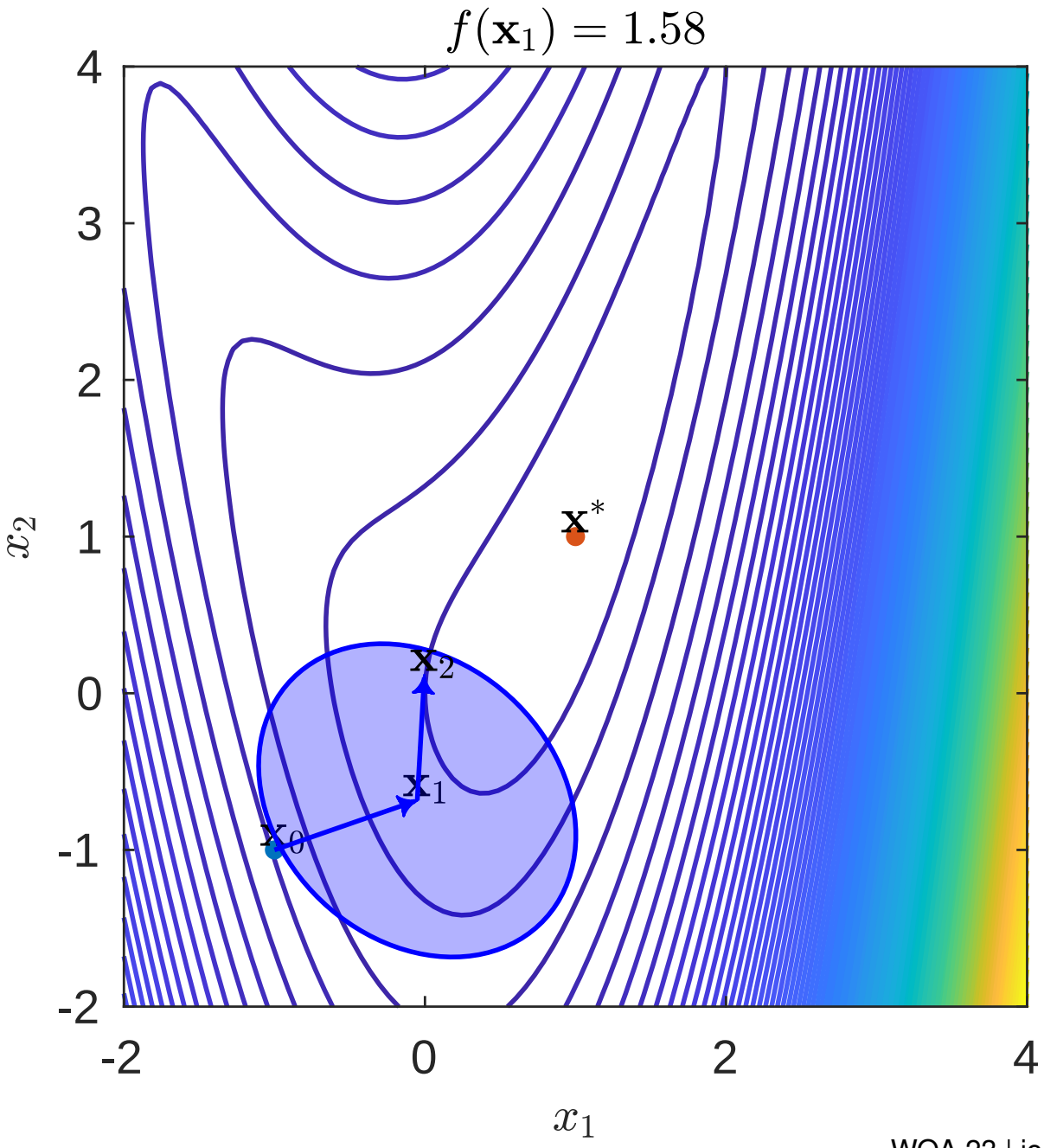
WTR-L2


$f(\mathbf{x}_1) = 1.58$
4
3
2
1
0
-1
-2
-2
0
2
4
x_2
x_1
$\mathbf{x}^*$
$\mathbf{x}_1$
$\mathbf{x}_0$

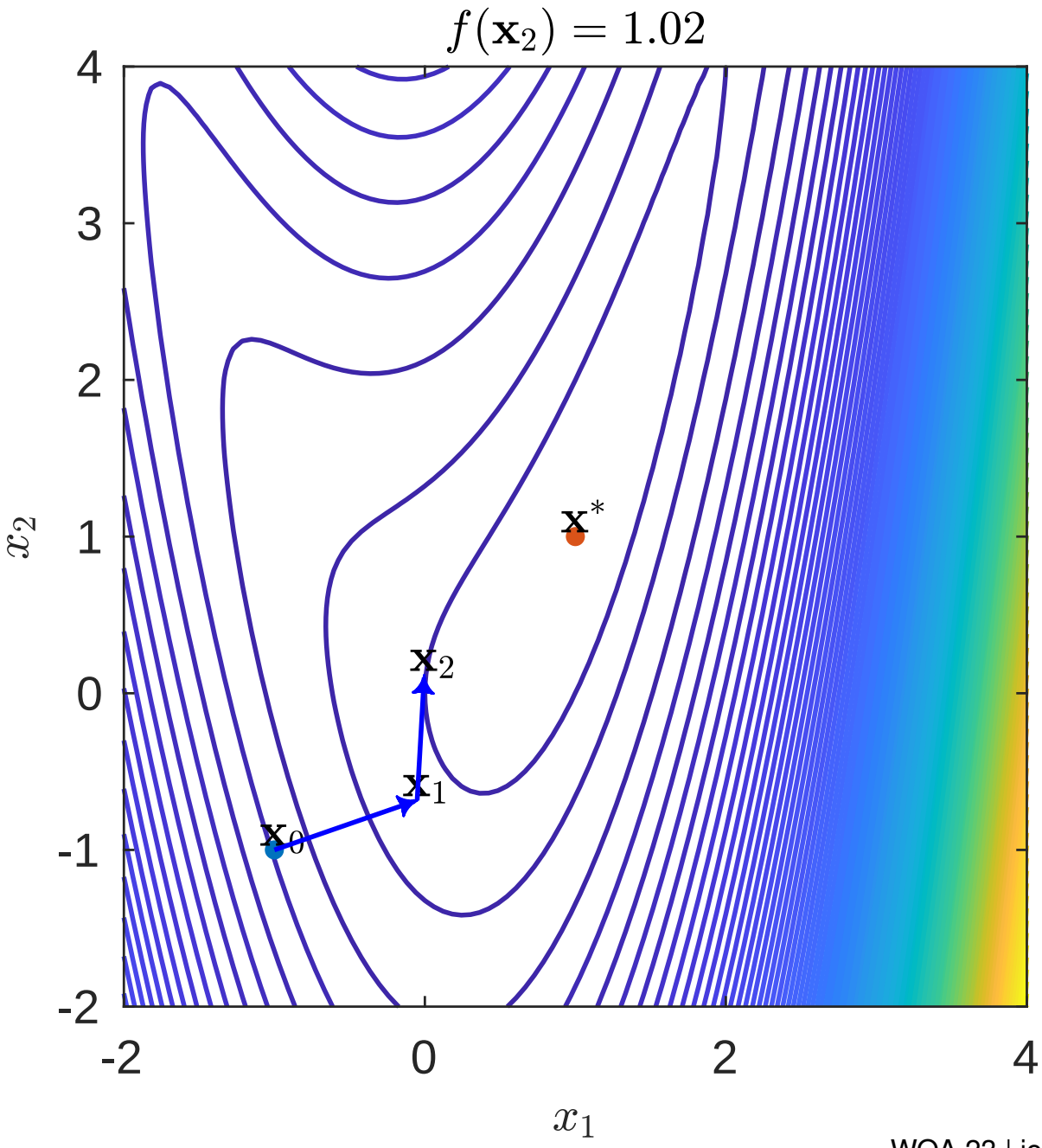
WTR-L2


$f(\mathbf{x}_1) = 1.58$
4
3
2
1
0
-1
-2
-2
0
2
4
x_2
x_1
$\mathbf{x}^*$
$\mathbf{x}_1$
$\mathbf{x}_0$

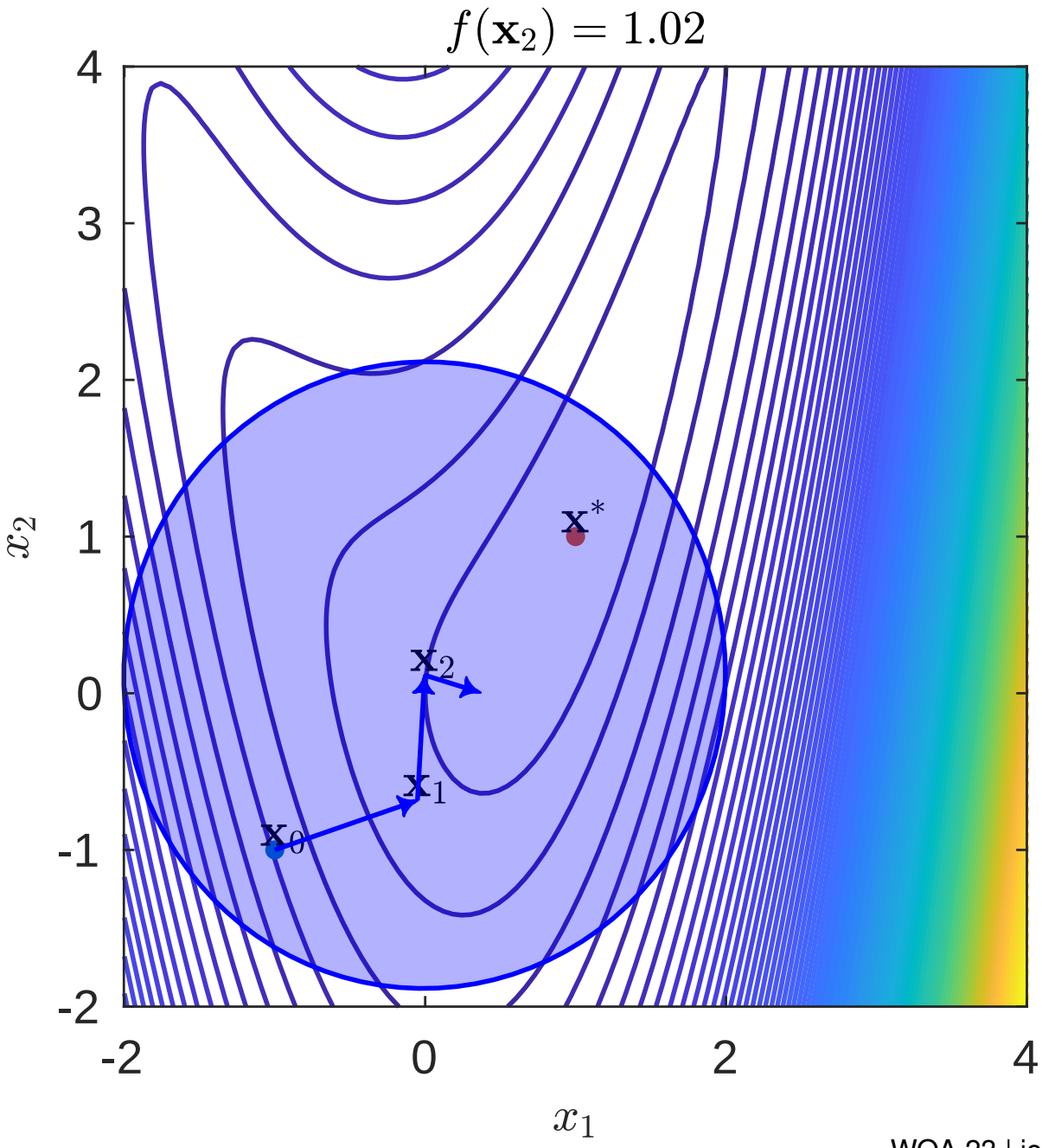
WTR-L2


$f(\mathbf{x}_1) = 1.58$
$\mathbf{x}^*$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$
x_2
x_1
4
3
2
1
0
-1
-2
-2
0
2
4

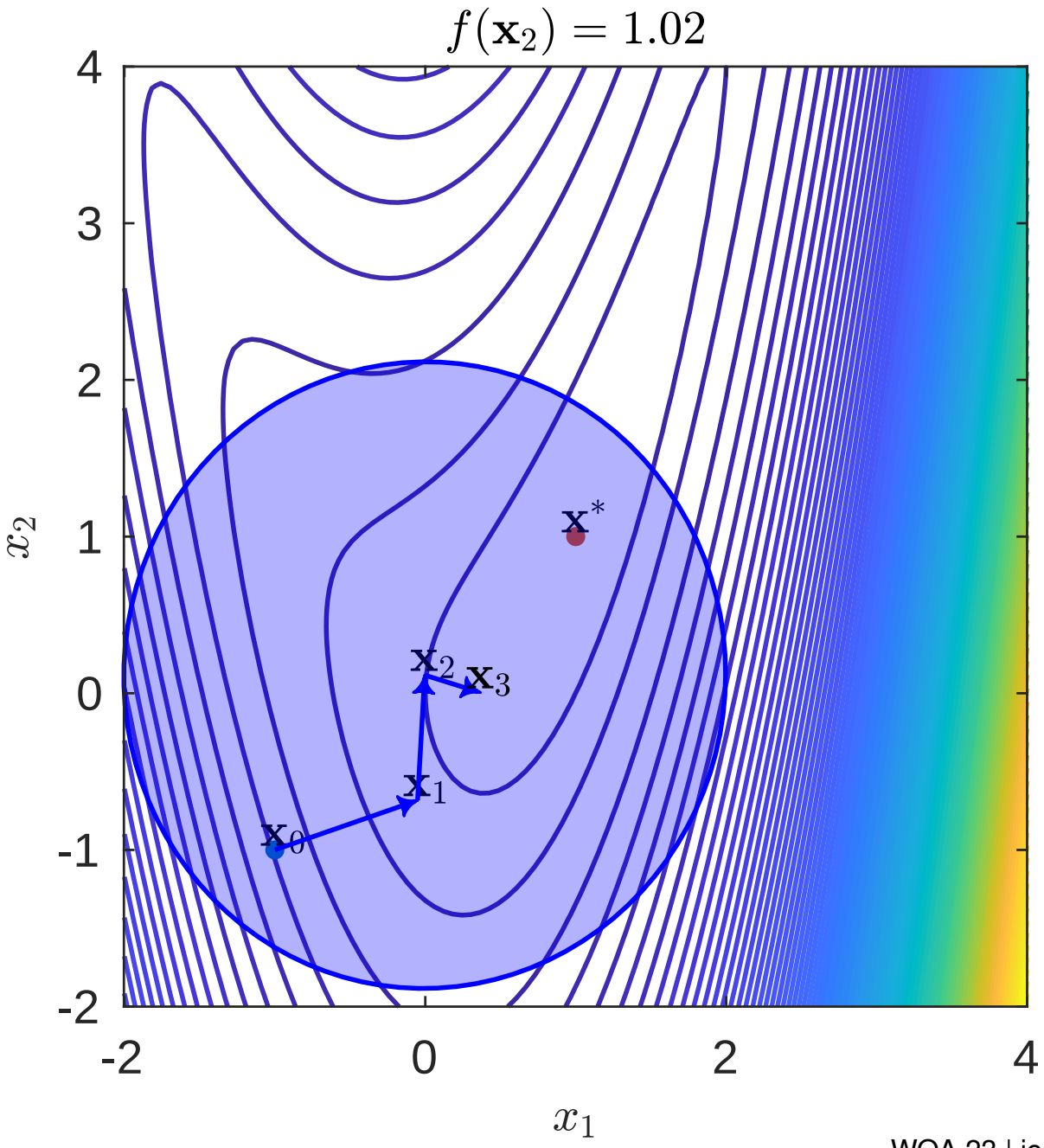
WTR-L2


$f(\mathbf{x}_2) = 1.02$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$

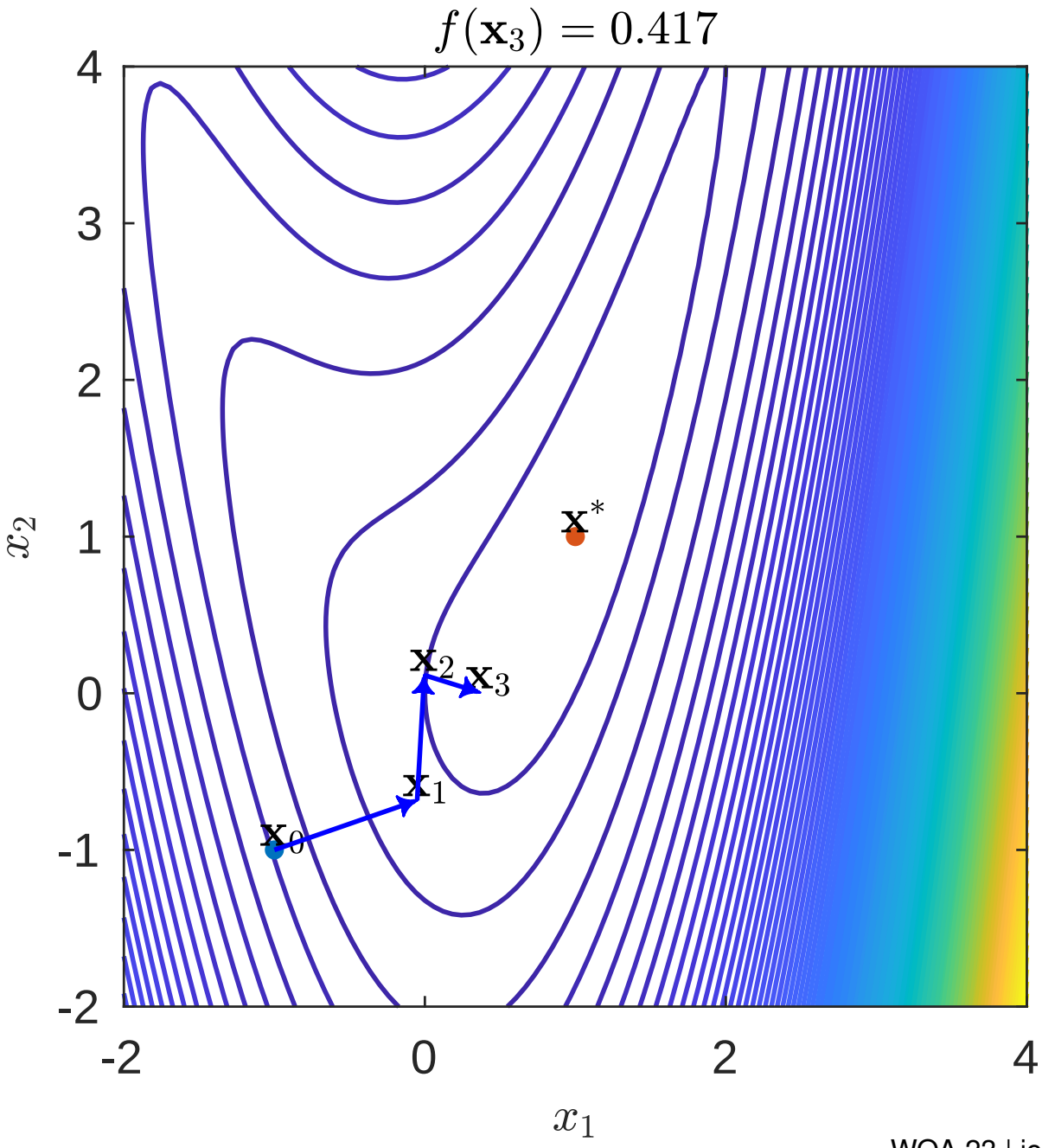
WTR-L2


$f(\mathbf{x}_2) = 1.02$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$

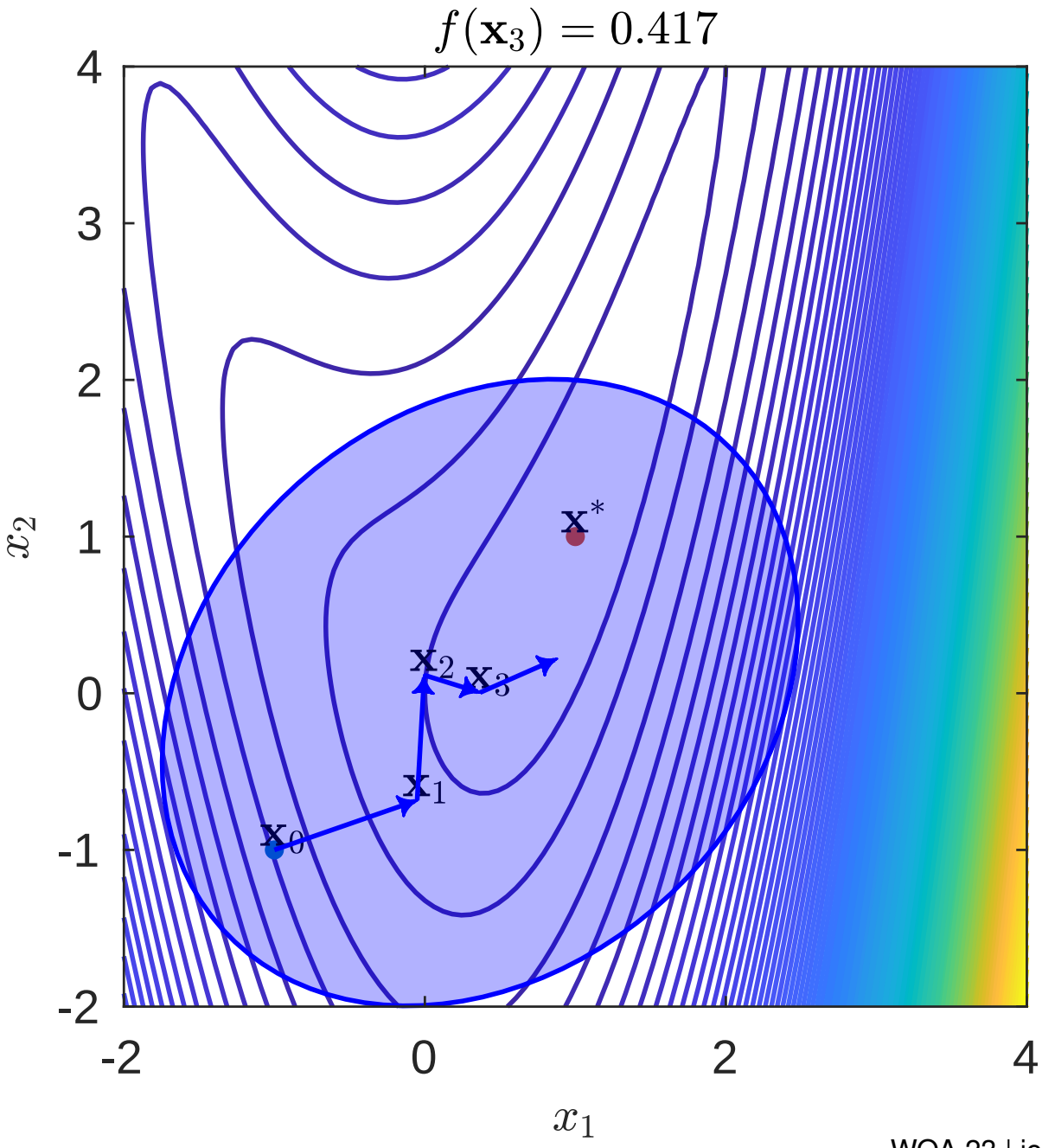
WTR-L2


$f(\mathbf{x}_2) = 1.02$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_1$
$\mathbf{x}_0$

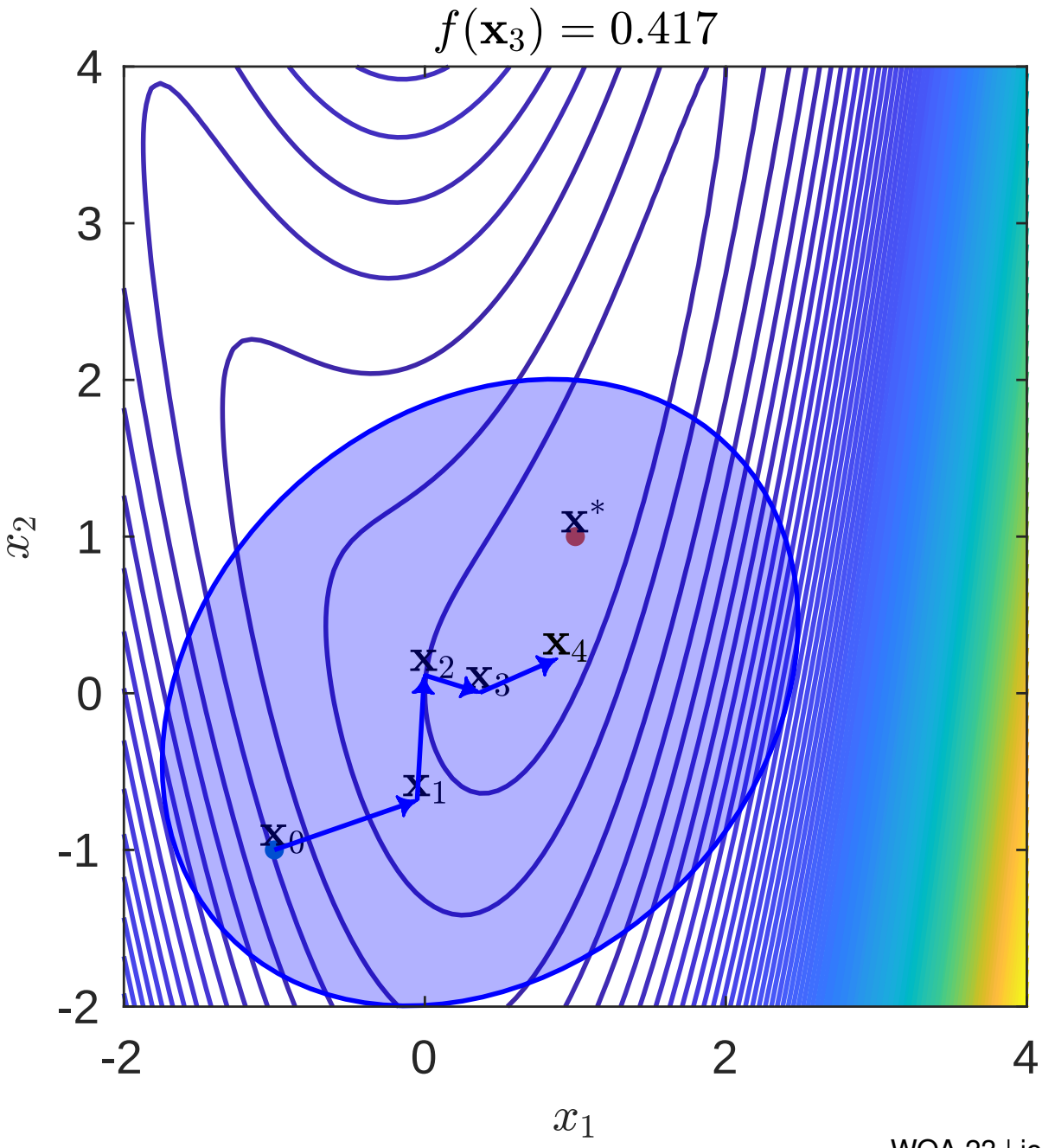
WTR-L2


$f(\mathbf{x}_3) = 0.417$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_1$
$\mathbf{x}_0$

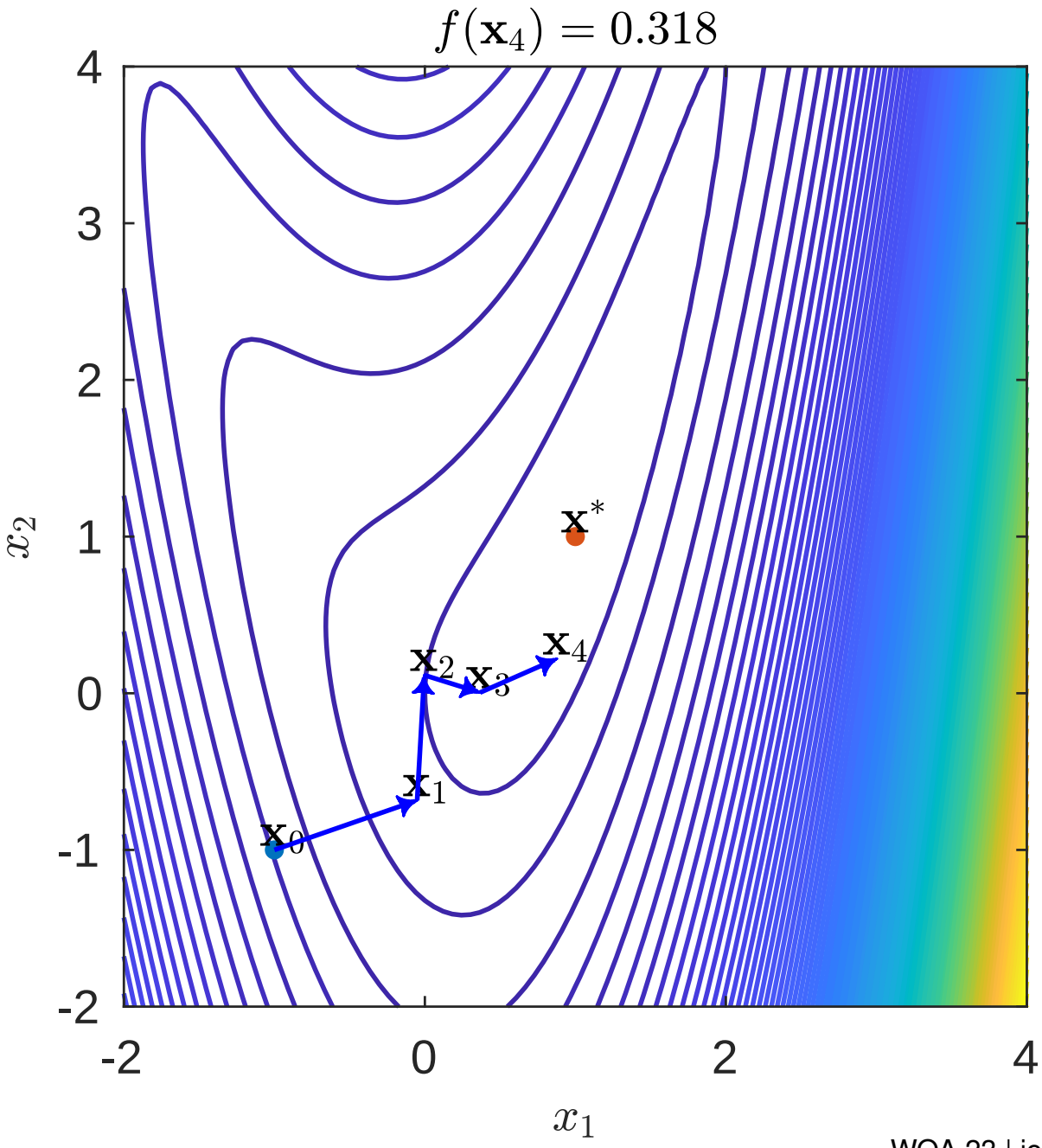
WTR-L2


$f(\mathbf{x}_3) = 0.417$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_1$
$\mathbf{x}_0$

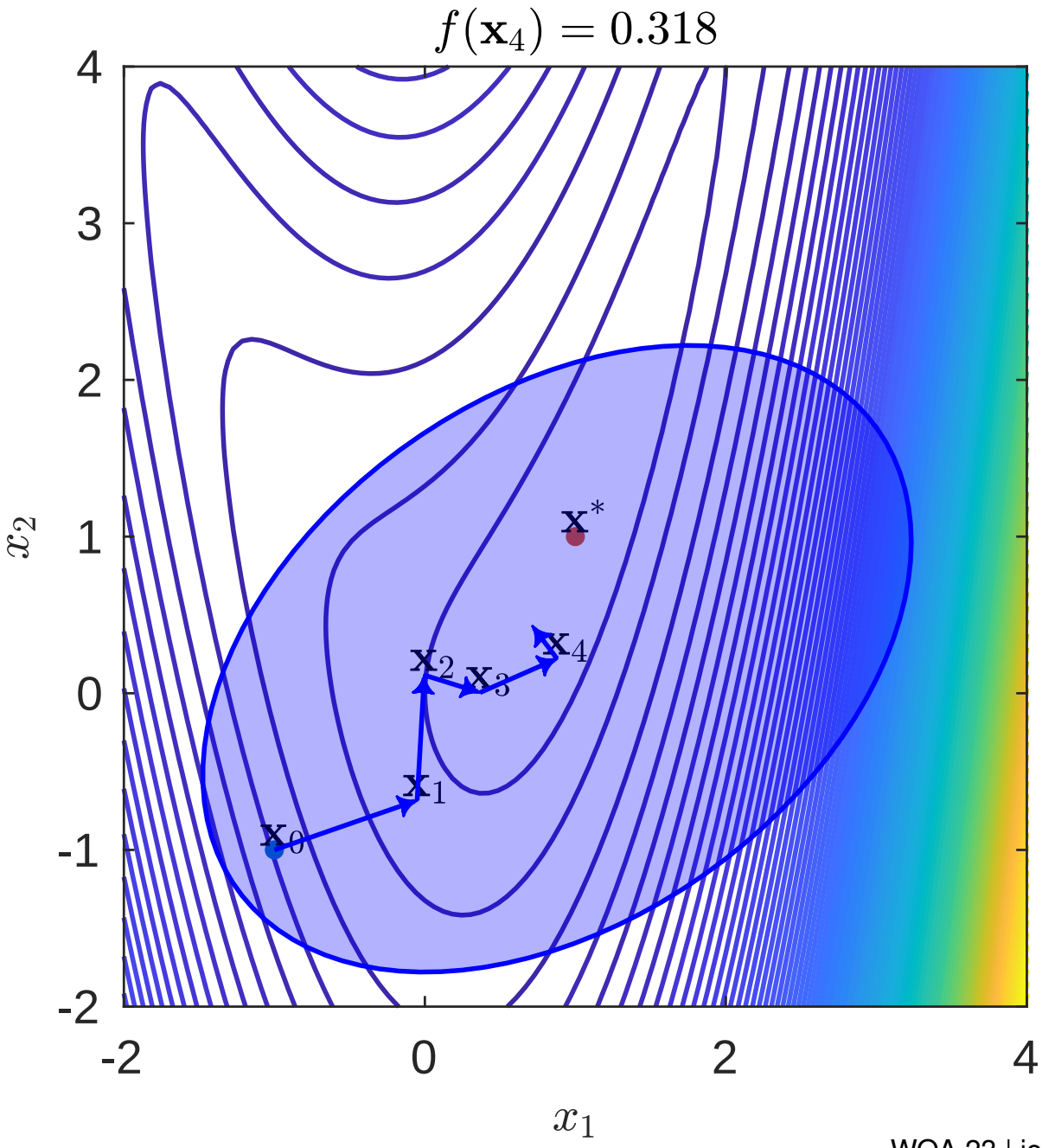
WTR-L2


$f(\mathbf{x}_3) = 0.417$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$

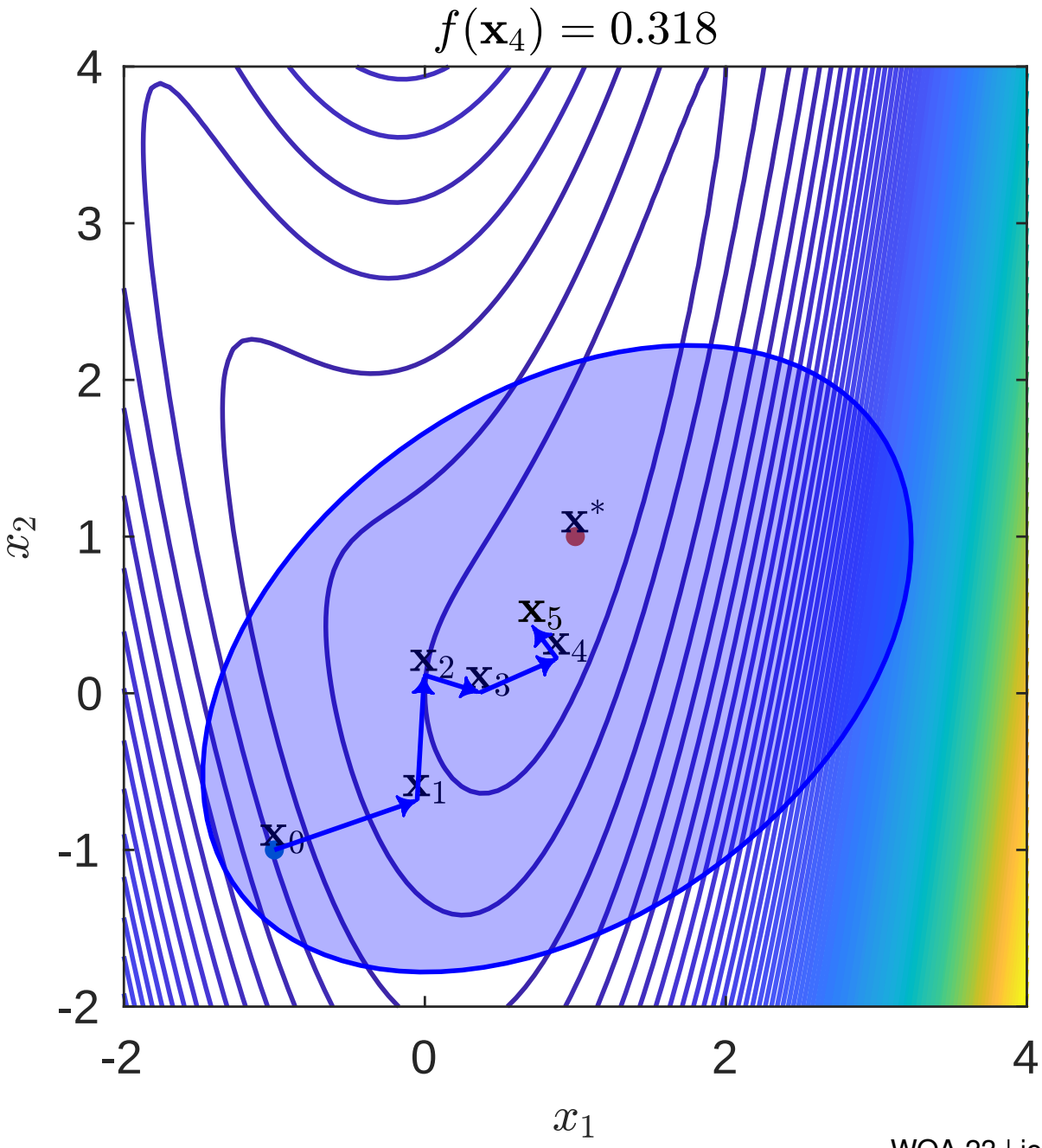
WTR-L2


$f(\mathbf{x}_4) = 0.318$
$\mathbf{x}^*$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

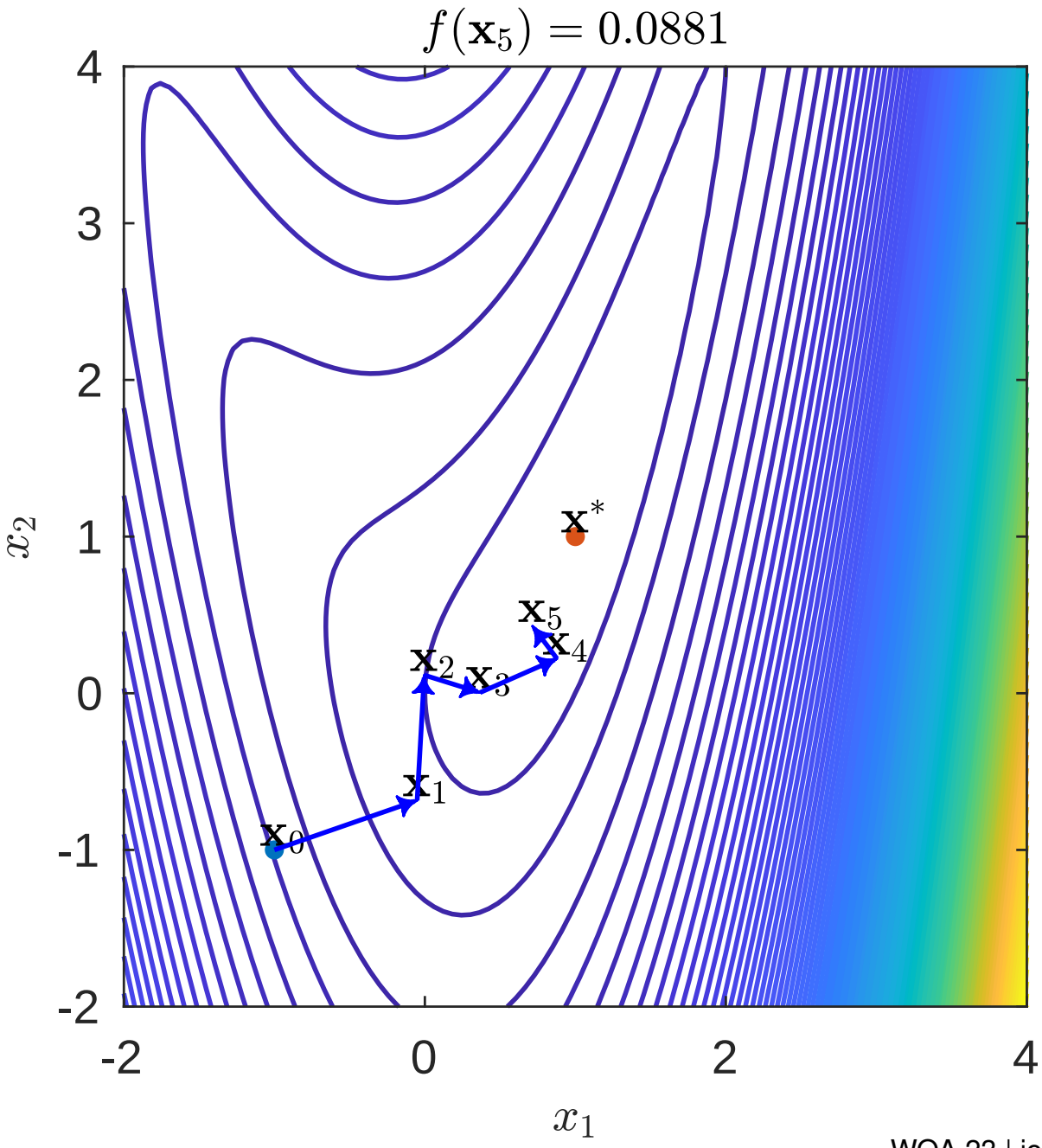
WTR-L2


$f(\mathbf{x}_4) = 0.318$
$\mathbf{x}^*$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
x_1
x_2
-2
0
2
4
-2
-1
0
1
2
3
4

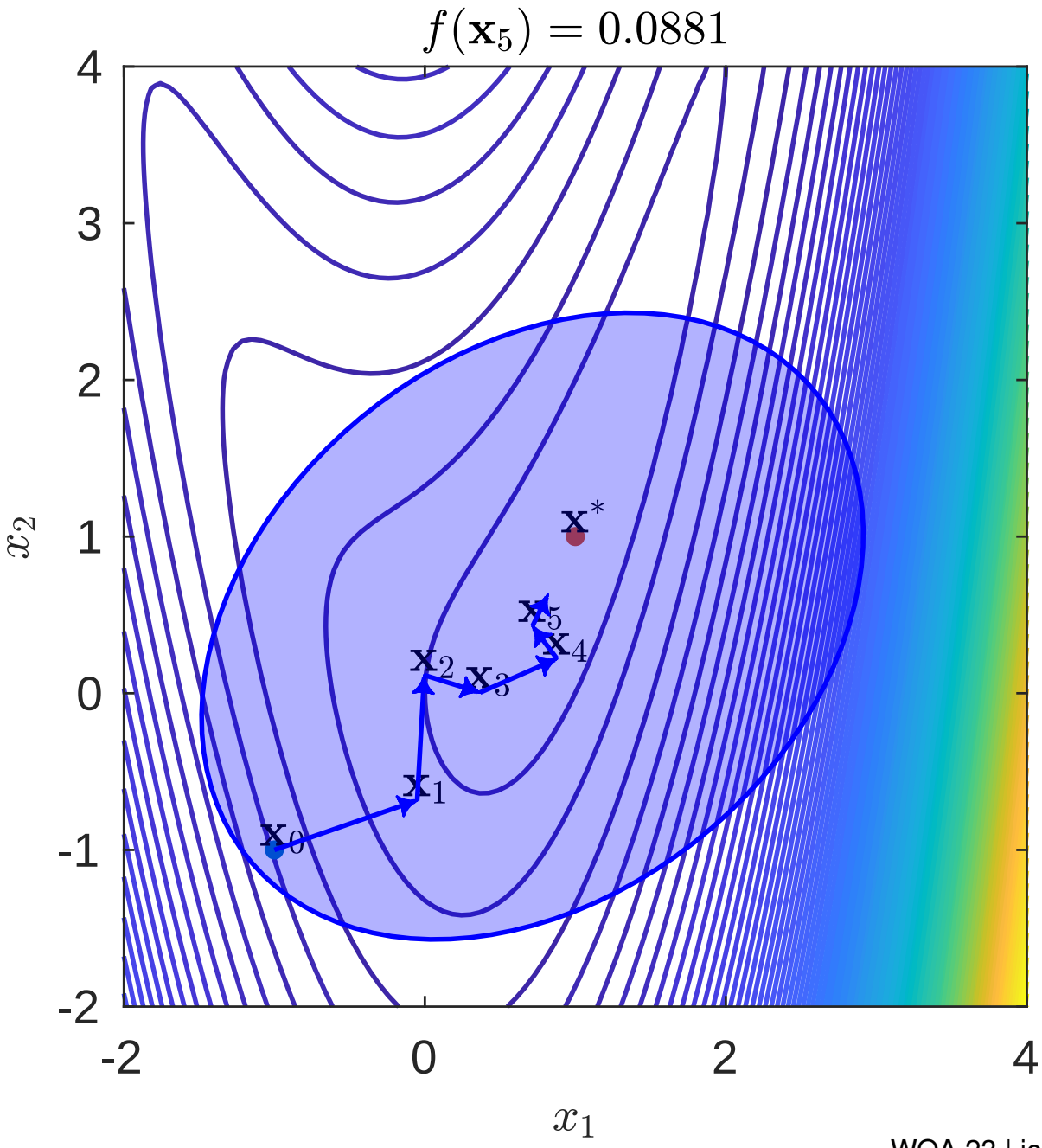
WTR-L2


$f(\mathbf{x}_4) = 0.318$
$\mathbf{x}^*$
$\mathbf{x}_5$
$\mathbf{x}_4$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_1$
$\mathbf{x}_0$
x_1
x_2
-2
0
2
4
-2
-1
0
1
2
3
4

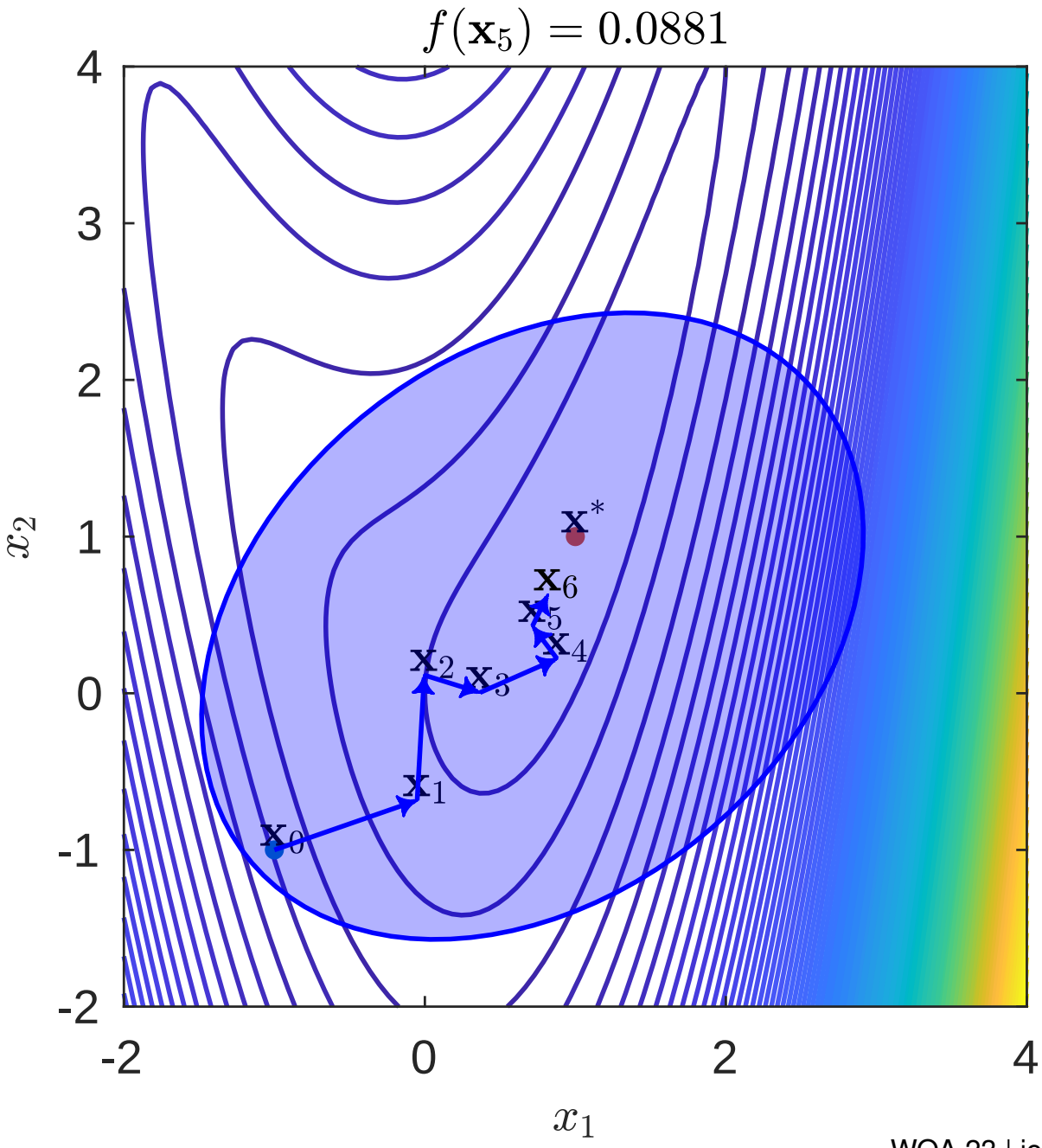
WTR-L2


$f(\mathbf{x}_5) = 0.0881$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_5$
$\mathbf{x}_4$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_1$
$\mathbf{x}_0$

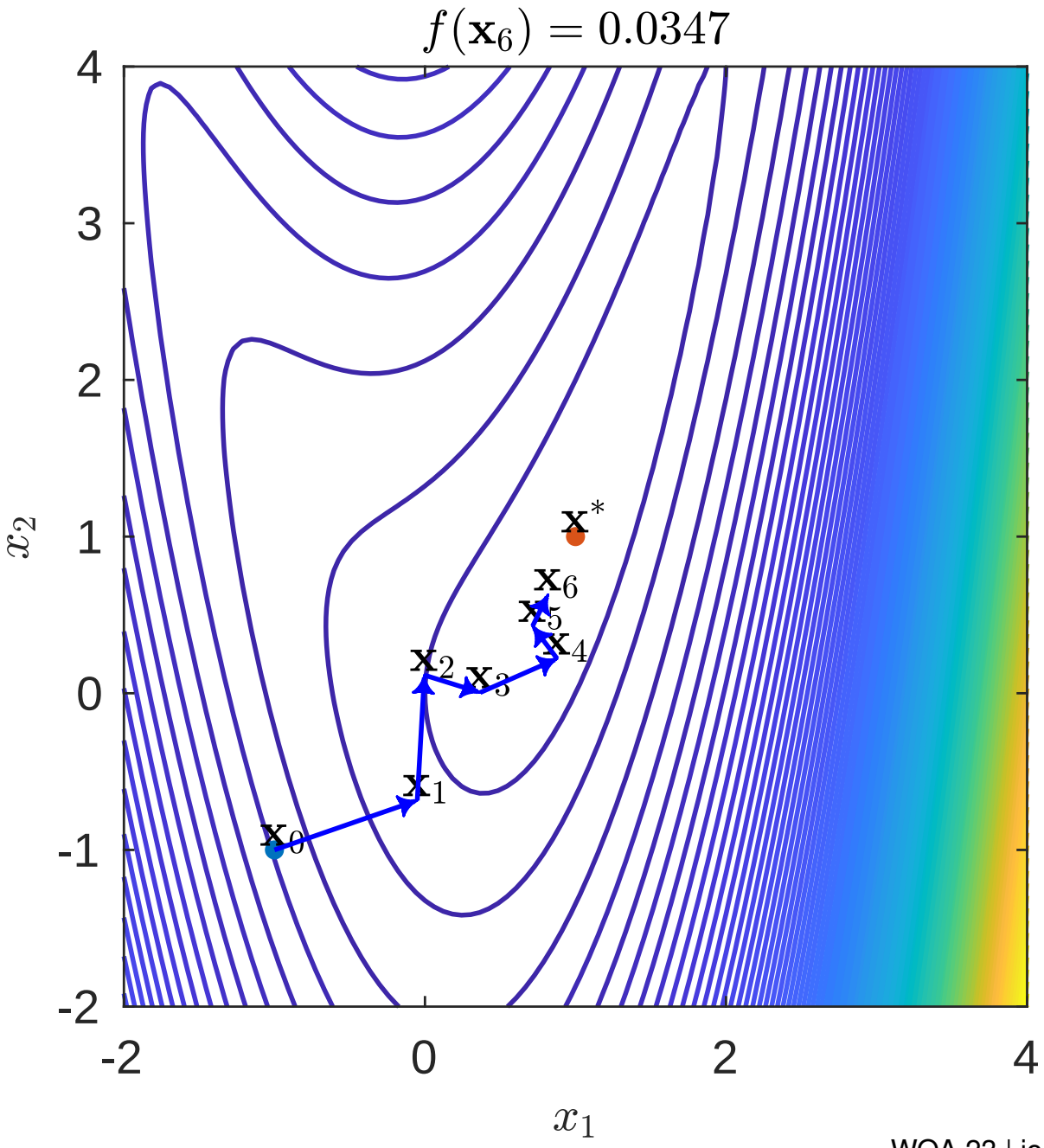
WTR-L2


$f(\mathbf{x}_5) = 0.0881$
$\mathbf{x}^*$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

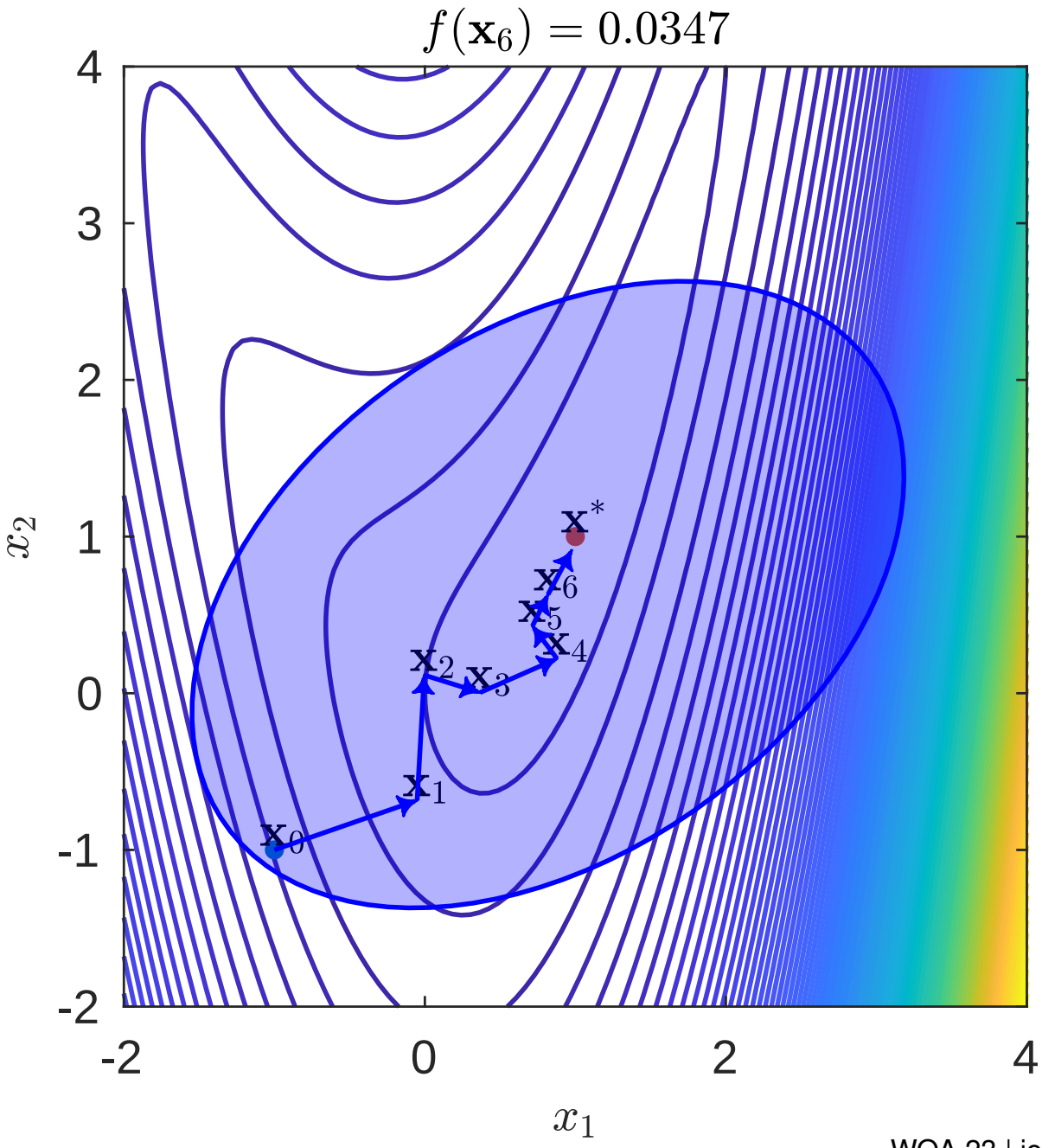
WTR-L2


$f(\mathbf{x}_5) = 0.0881$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}^*$
$\mathbf{x}_6$
$\mathbf{x}_5$
$\mathbf{x}_4$
$\mathbf{x}_3$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$

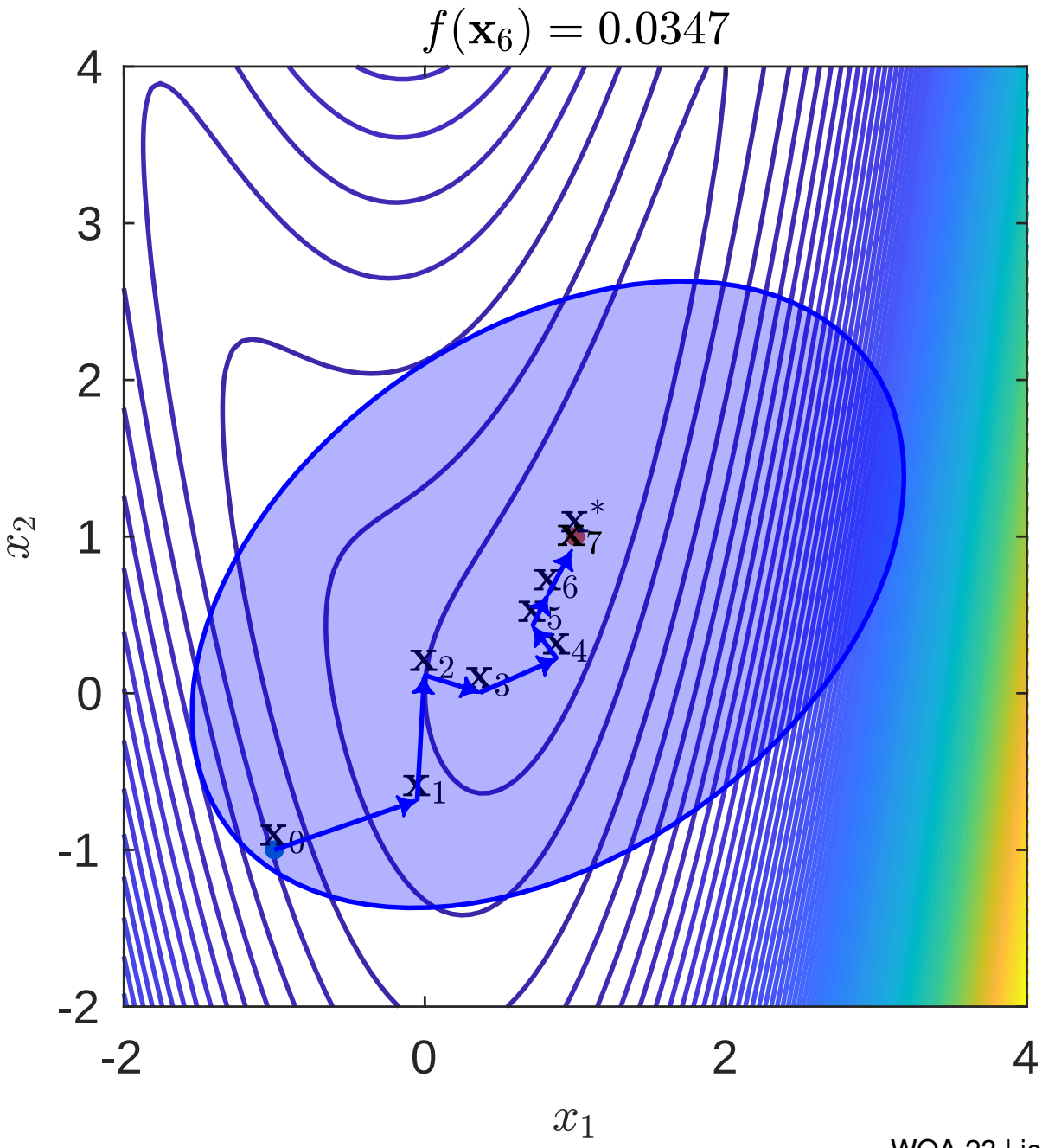
WTR-L2


$f(\mathbf{x}_6) = 0.0347$
$\mathbf{x}^*$
$\mathbf{x}_6$
$\mathbf{x}_5$
$\mathbf{x}_4$
$\mathbf{x}_3$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

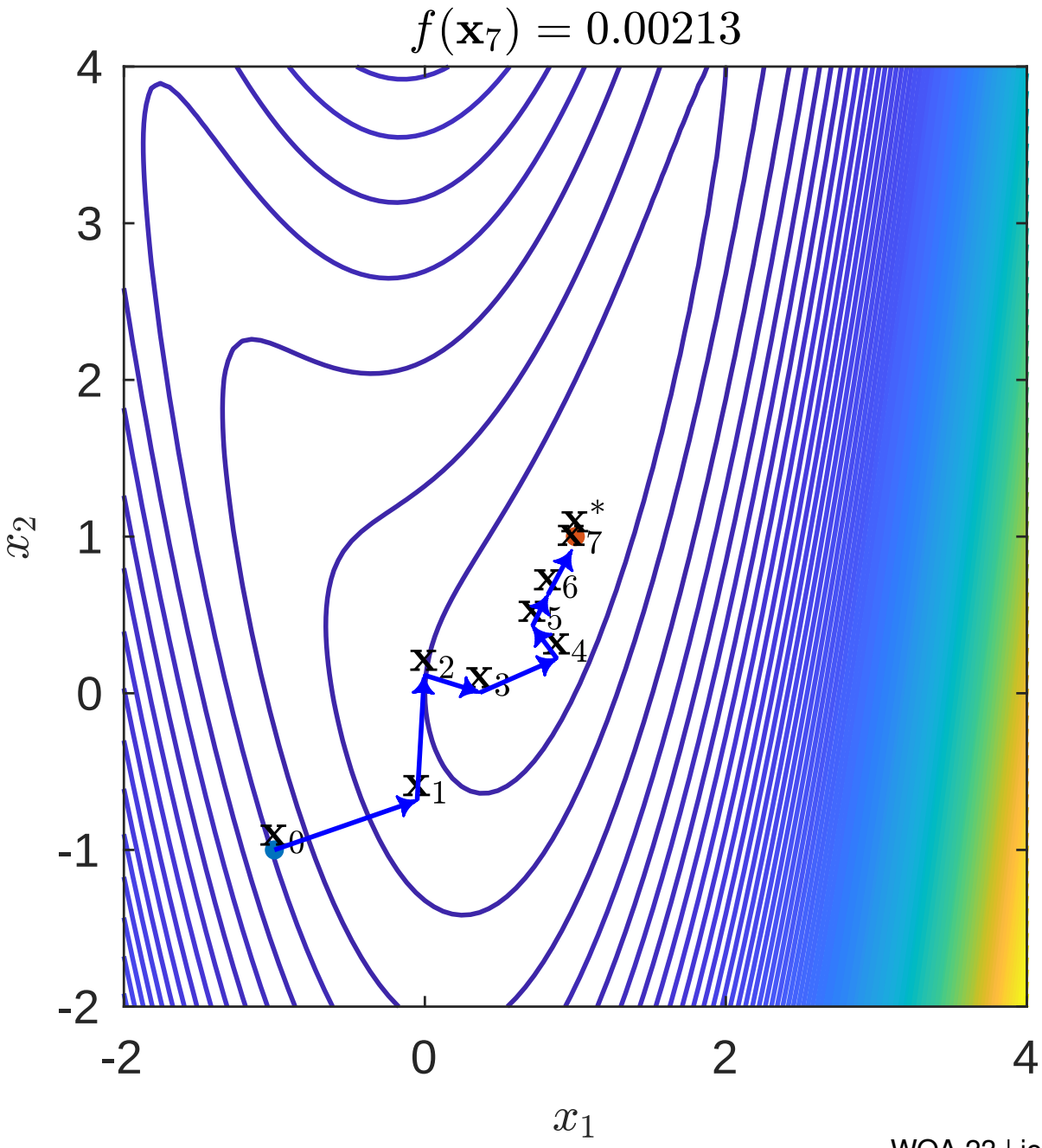
WTR-L2


$f(\mathbf{x}_6) = 0.0347$
$\mathbf{x}^*$
$\mathbf{x}_6$
$\mathbf{x}_5$
$\mathbf{x}_4$
$\mathbf{x}_3$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

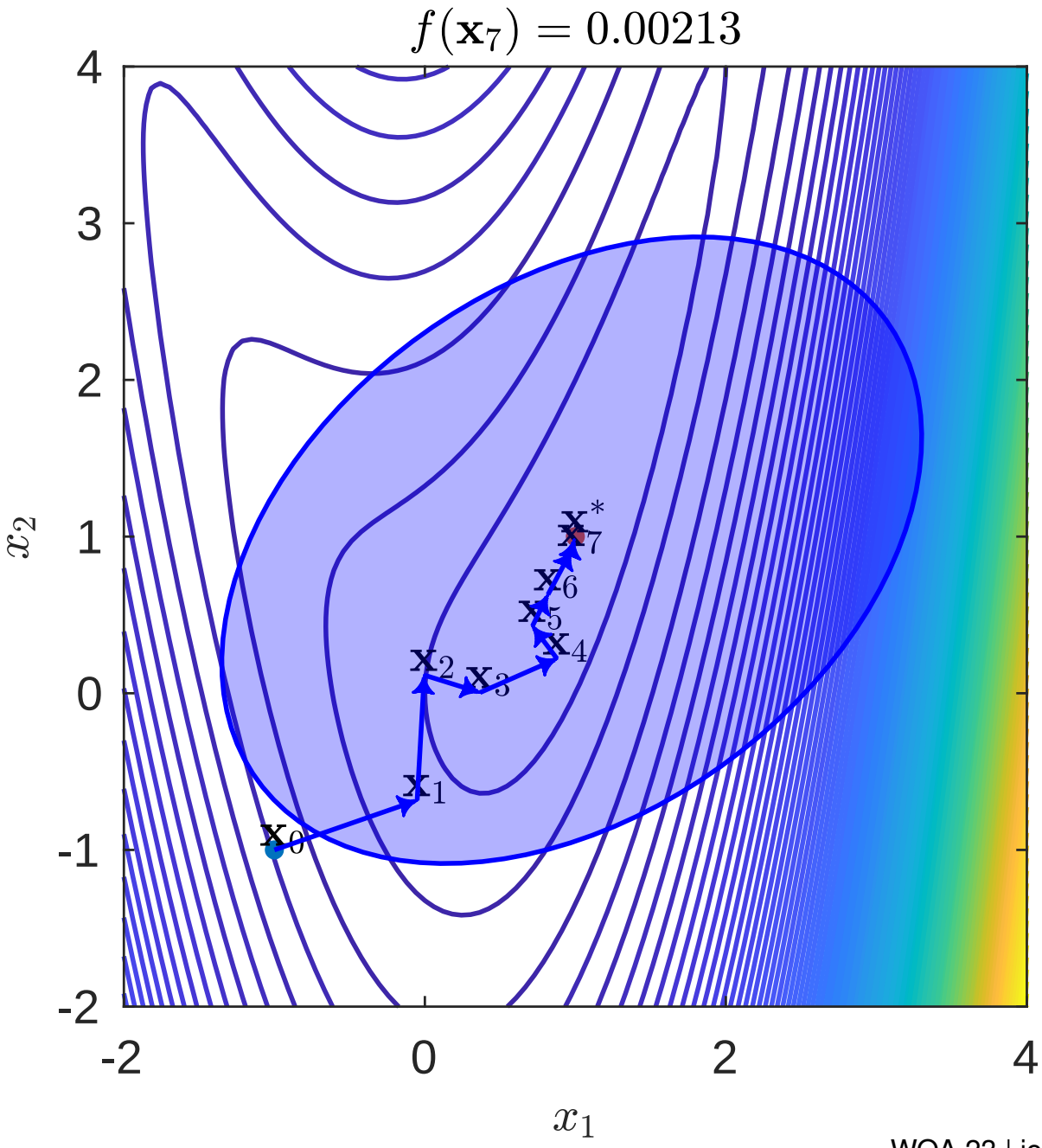
WTR-L2


$f(\mathbf{x}_6) = 0.0347$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}^*$

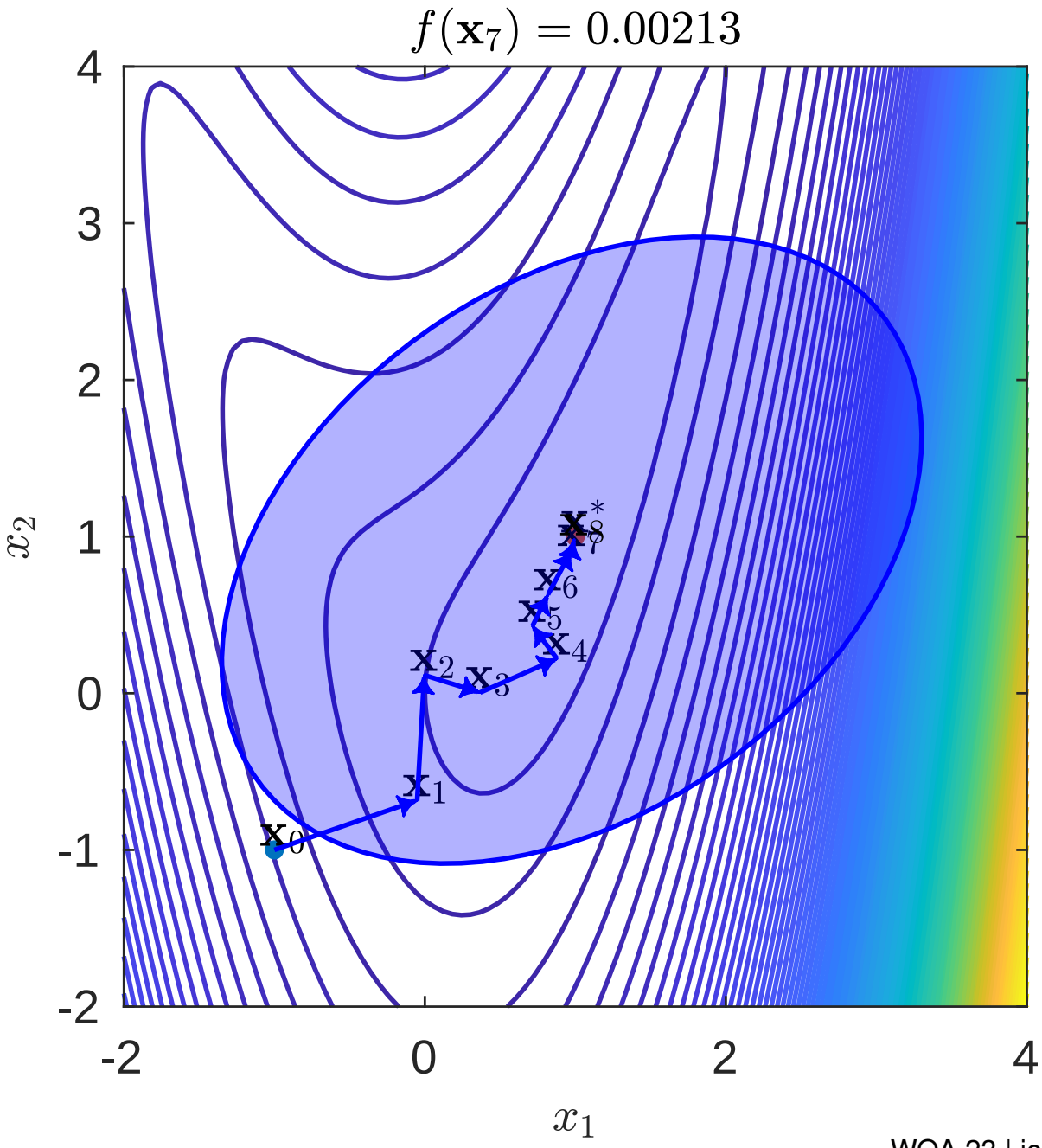
WTR-L2


$f(\mathbf{x}_7) = 0.00213$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}^*$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

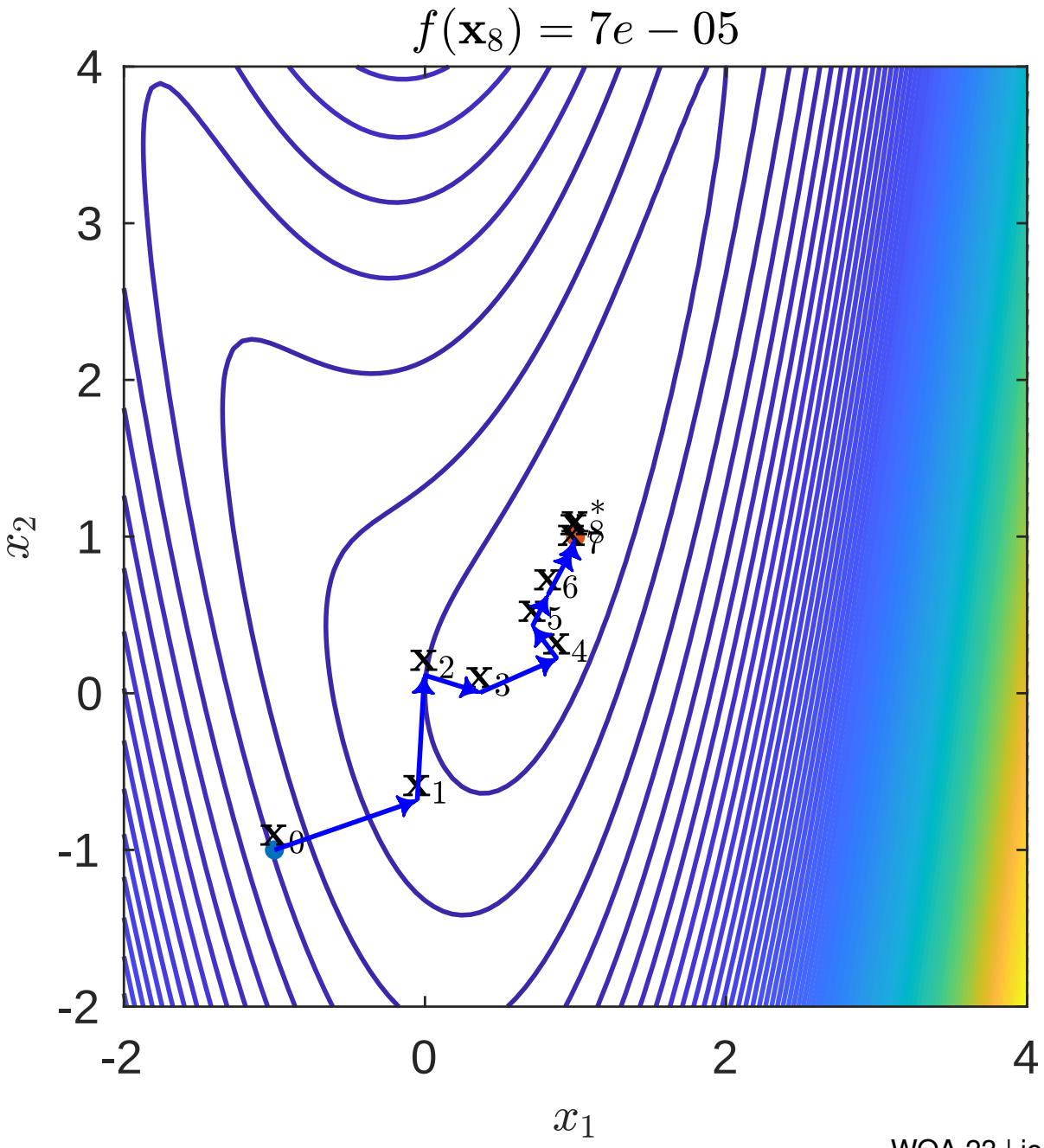
WTR-L2


$f(\mathbf{x}_7) = 0.00213$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}^*$

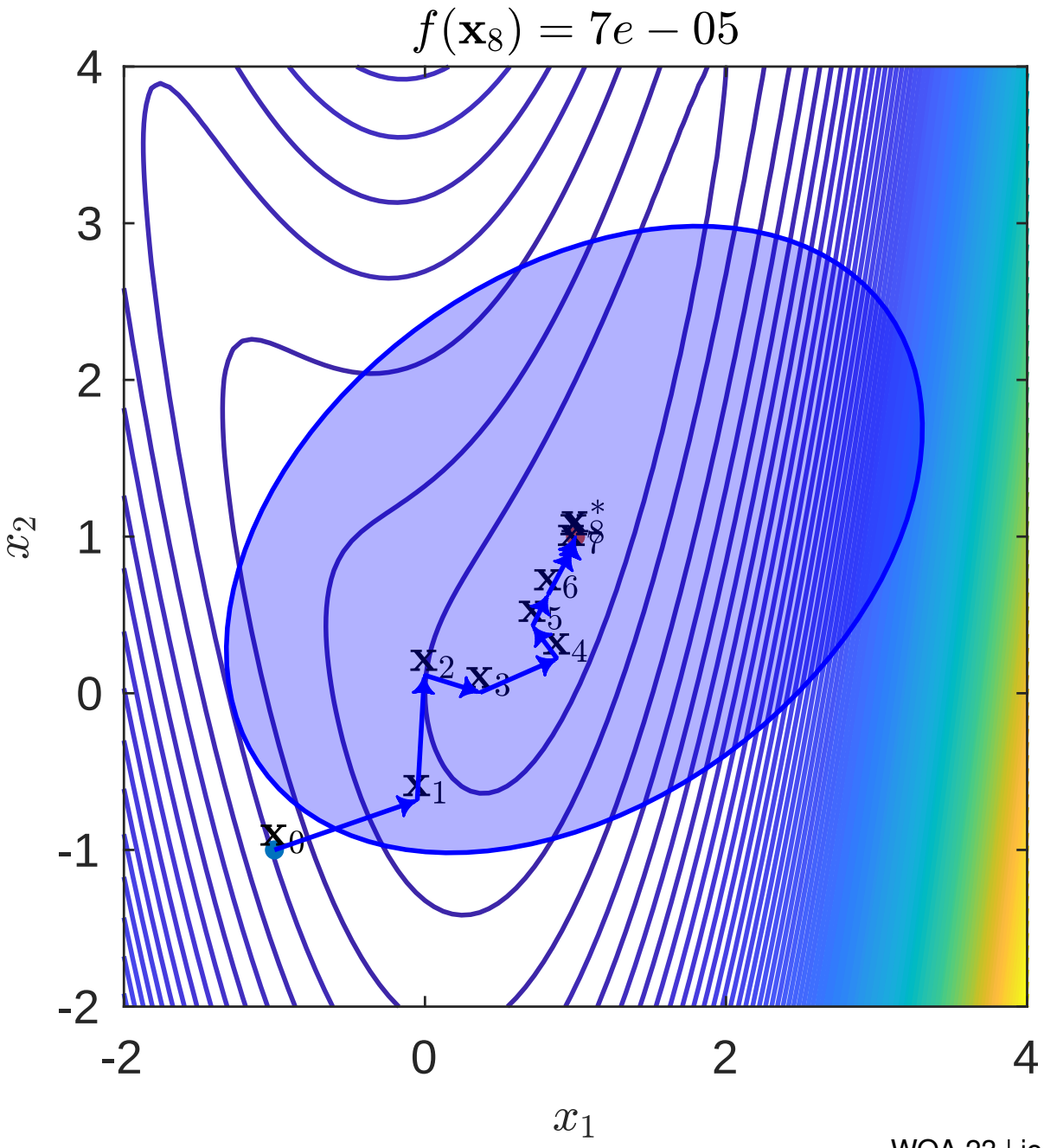
WTR-L2


$f(\mathbf{x}_7) = 0.00213$
4
3
2
1
0
-1
-2
-2
0
2
4
x_1
x_2
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}_8$
$\mathbf{x}_8^*$

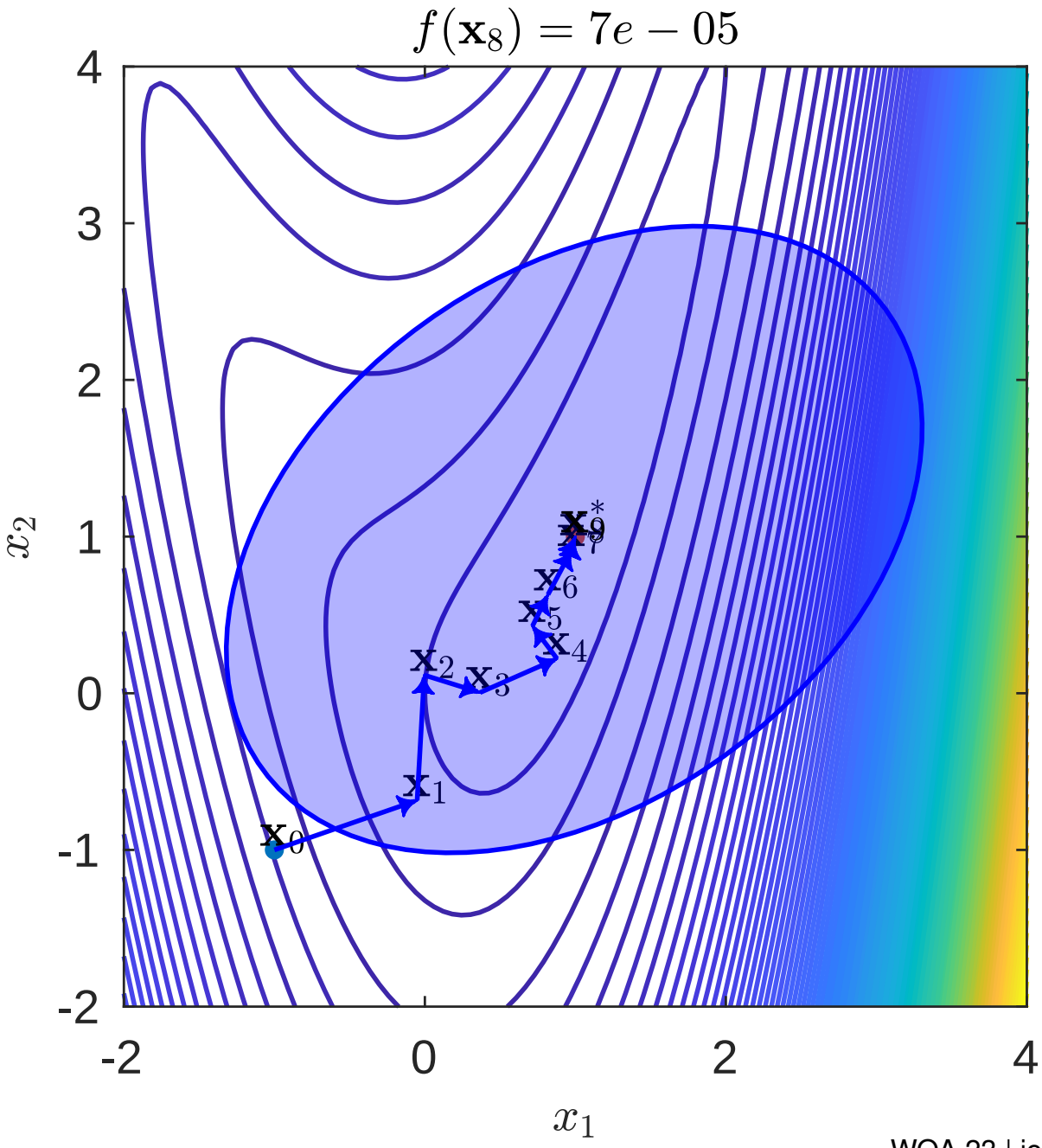
WTR-L2


$f(\mathbf{x}_8) = 7e-05$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}_8$
$\mathbf{x}^*$
x_1
x_2
-2
0
2
4
-2
-1
0
1
2
3
4

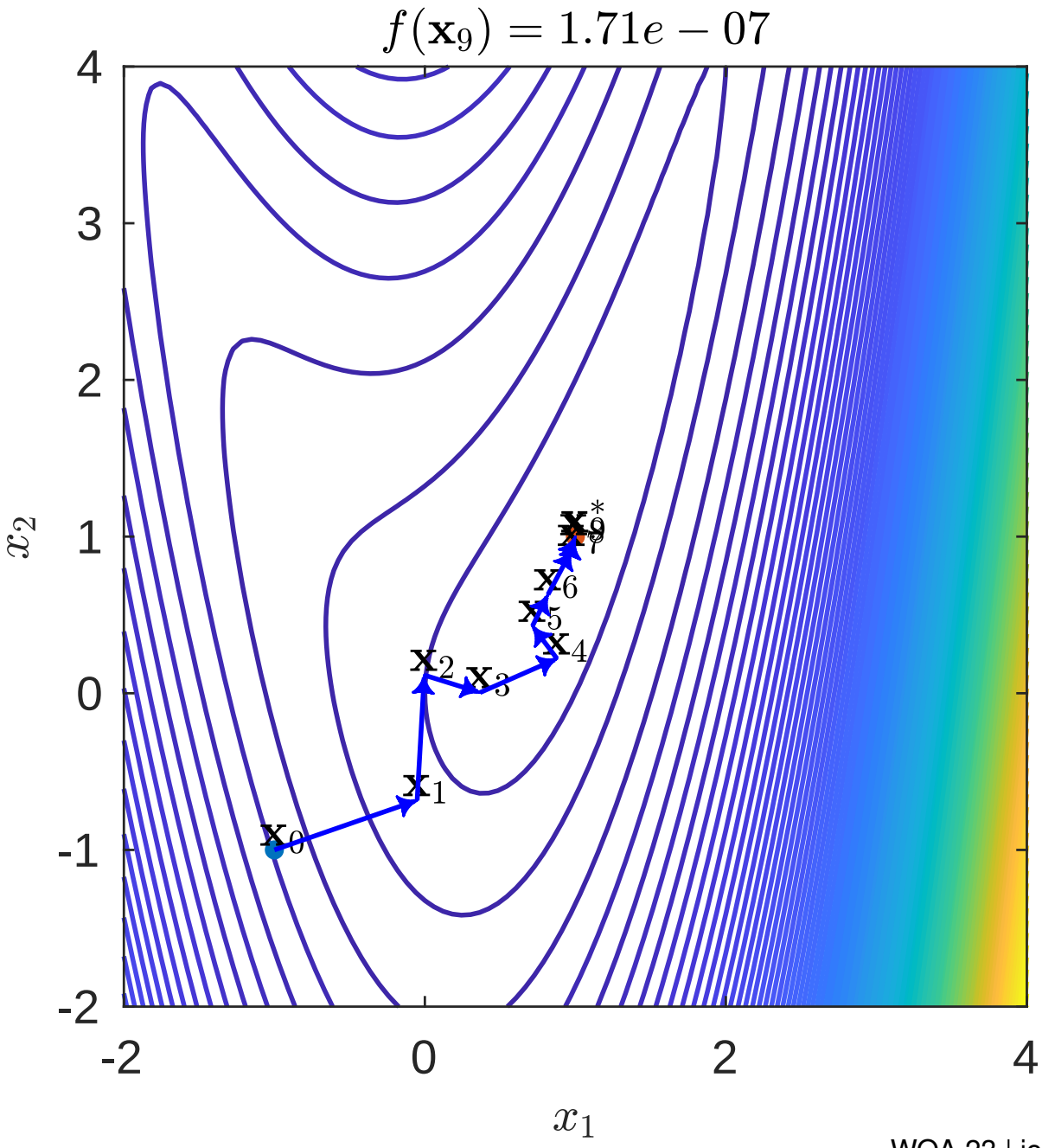
WTR-L2


$f(\mathbf{x}_8) = 7e-05$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}_8$
$\mathbf{x}_8^*$
x_1
x_2
-2
0
2
4
-2
-1
0
1
2
3
4

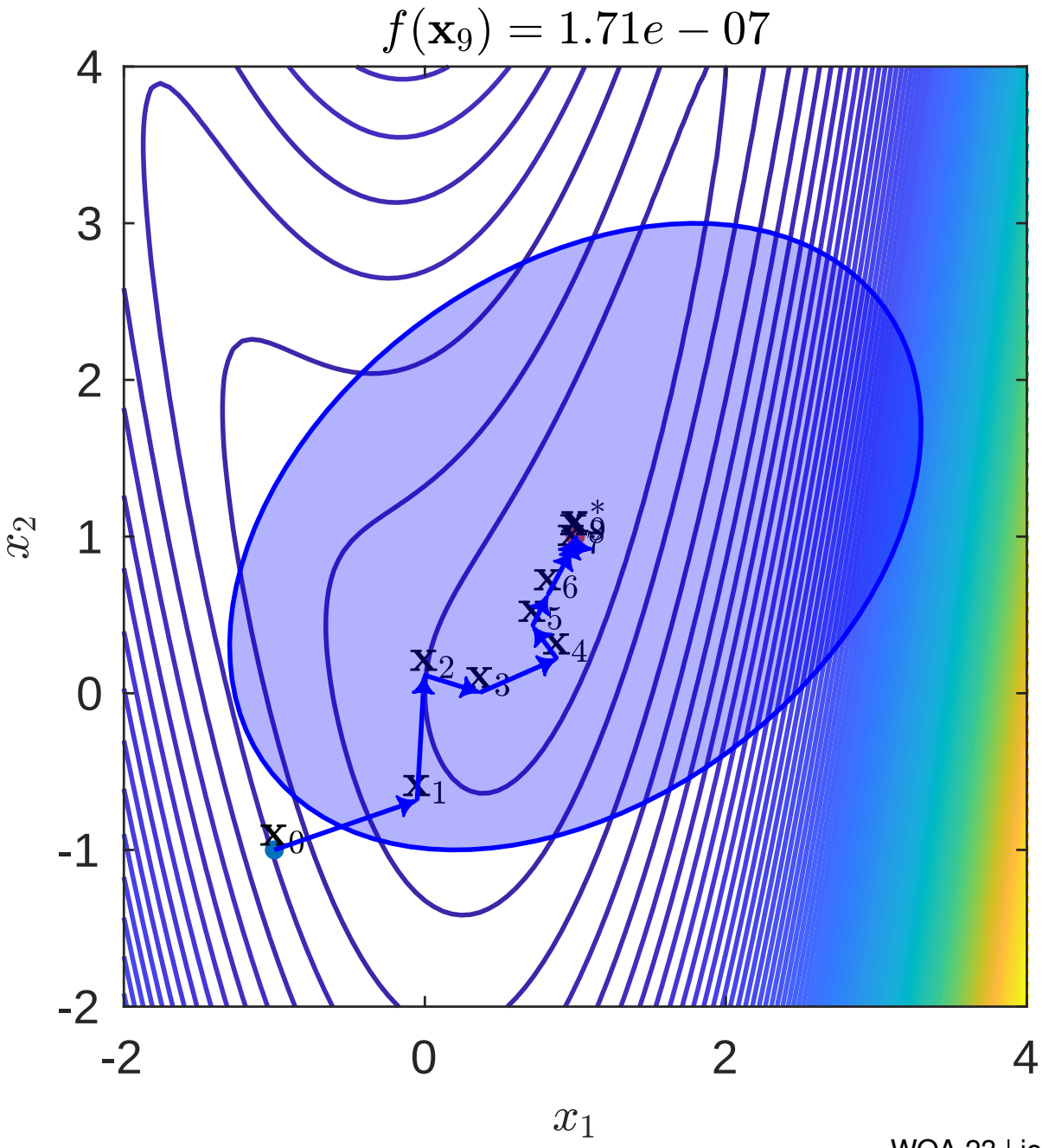
WTR-L2


$f(\mathbf{x}_8) = 7e-05$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}_9^*$
x_1
x_2
-2
0
2
4
-2
-1
0
1
2
3
4

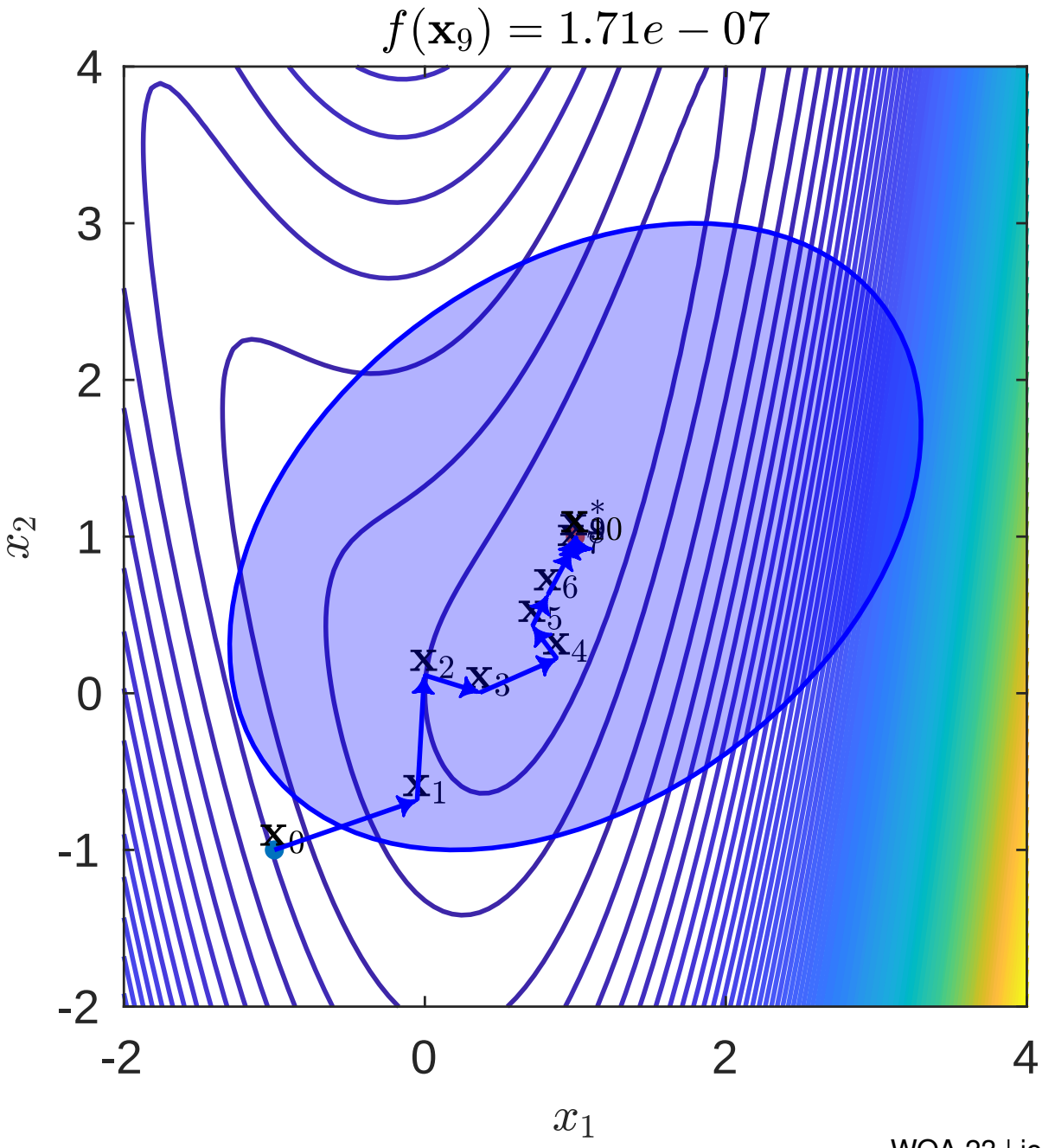
WTR-L2


$f(\mathbf{x}_9) = 1.71e-07$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_7$
$\mathbf{x}_9$
$\mathbf{x}_9^*$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

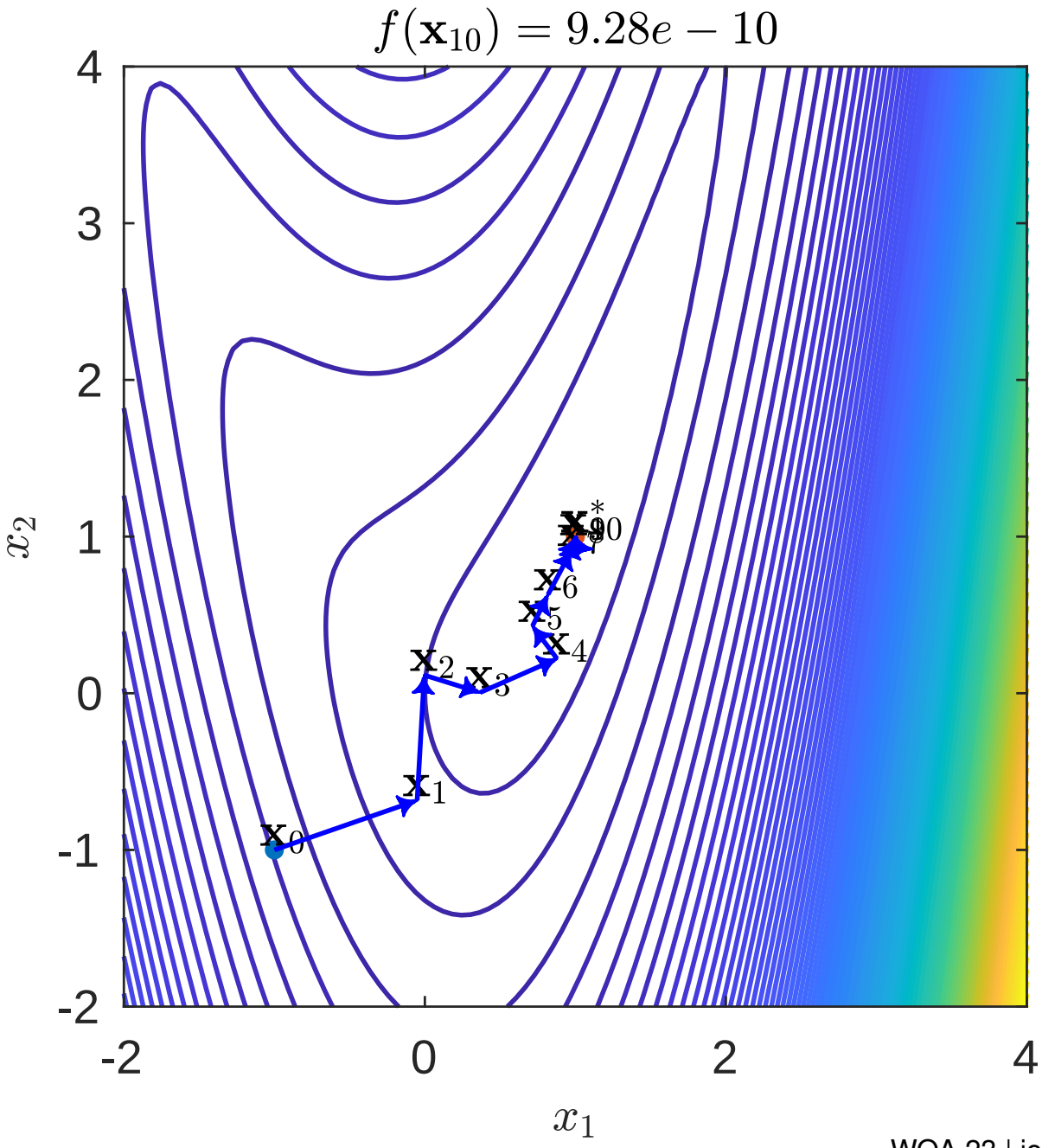
WTR-L2


$f(\mathbf{x}_9) = 1.71e-07$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_9$
$\mathbf{x}_9^*$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

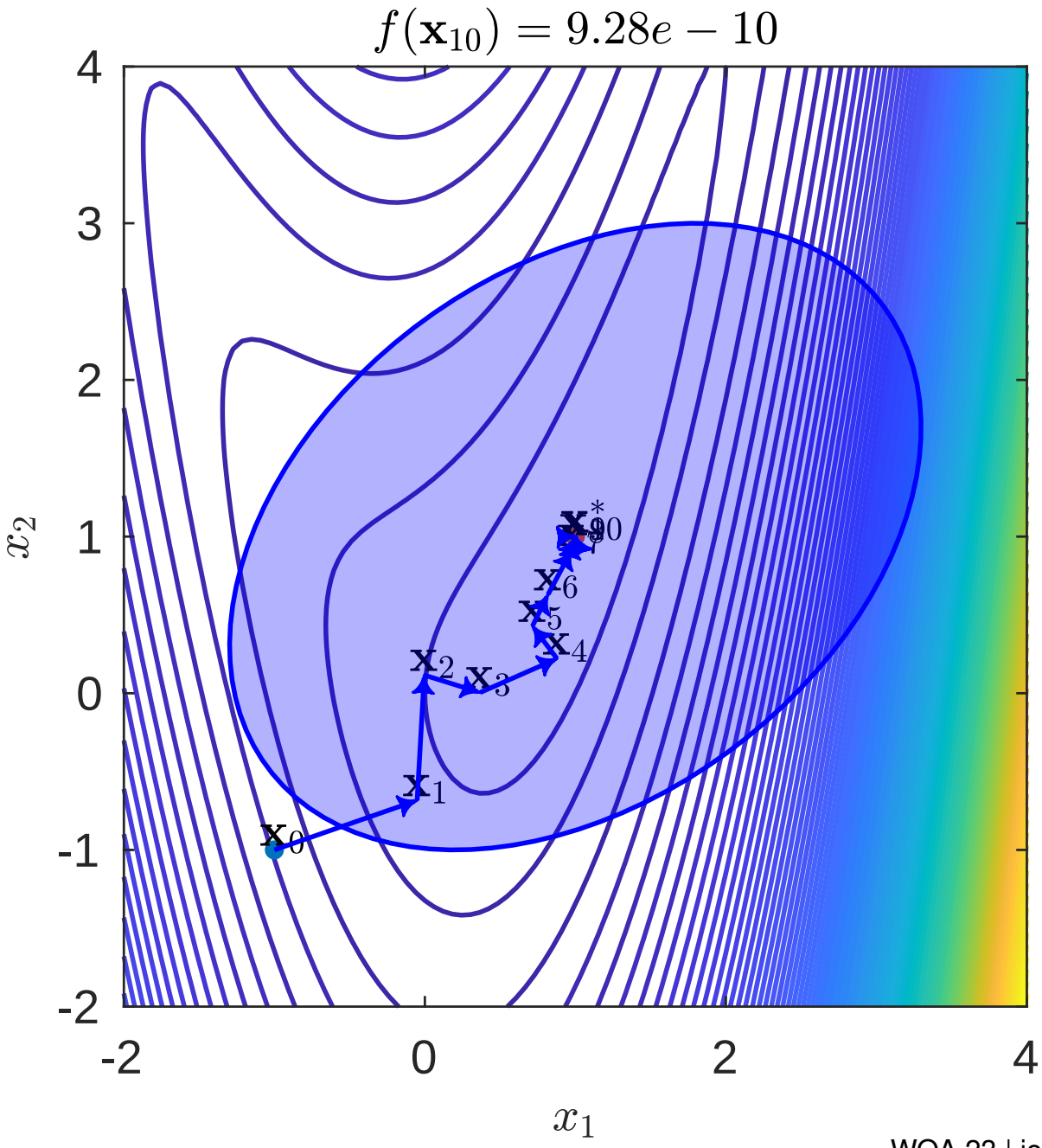
WTR-L2


$f(\mathbf{x}_9) = 1.71e-07$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_9$
$\mathbf{x}_{10}^*$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

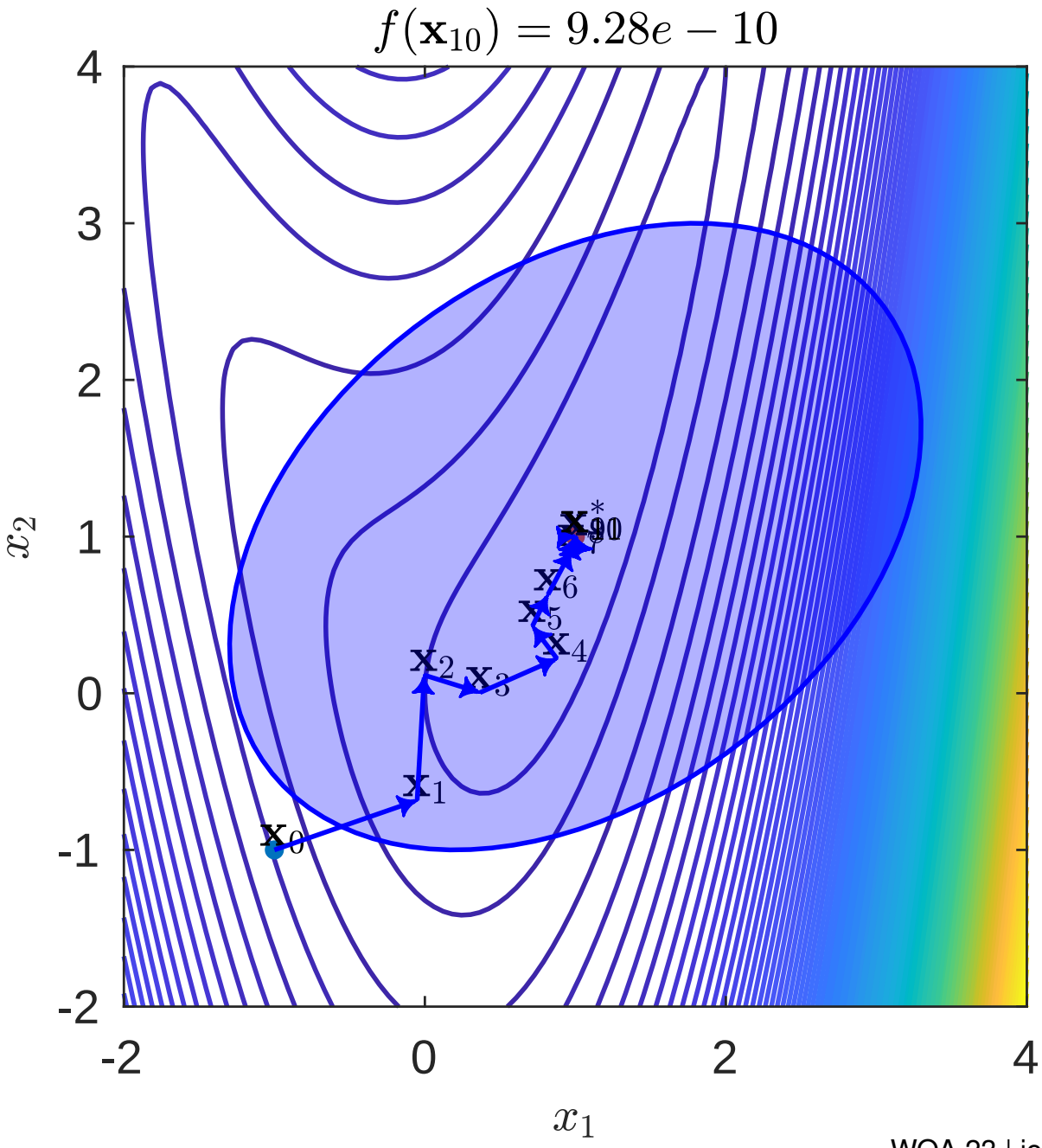
WTR-L2


$f(\mathbf{x}_{10}) = 9.28e - 10$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_{10}$
$\mathbf{x}^*$
x_1
x_2
-2
0
2
4
-1
1
3

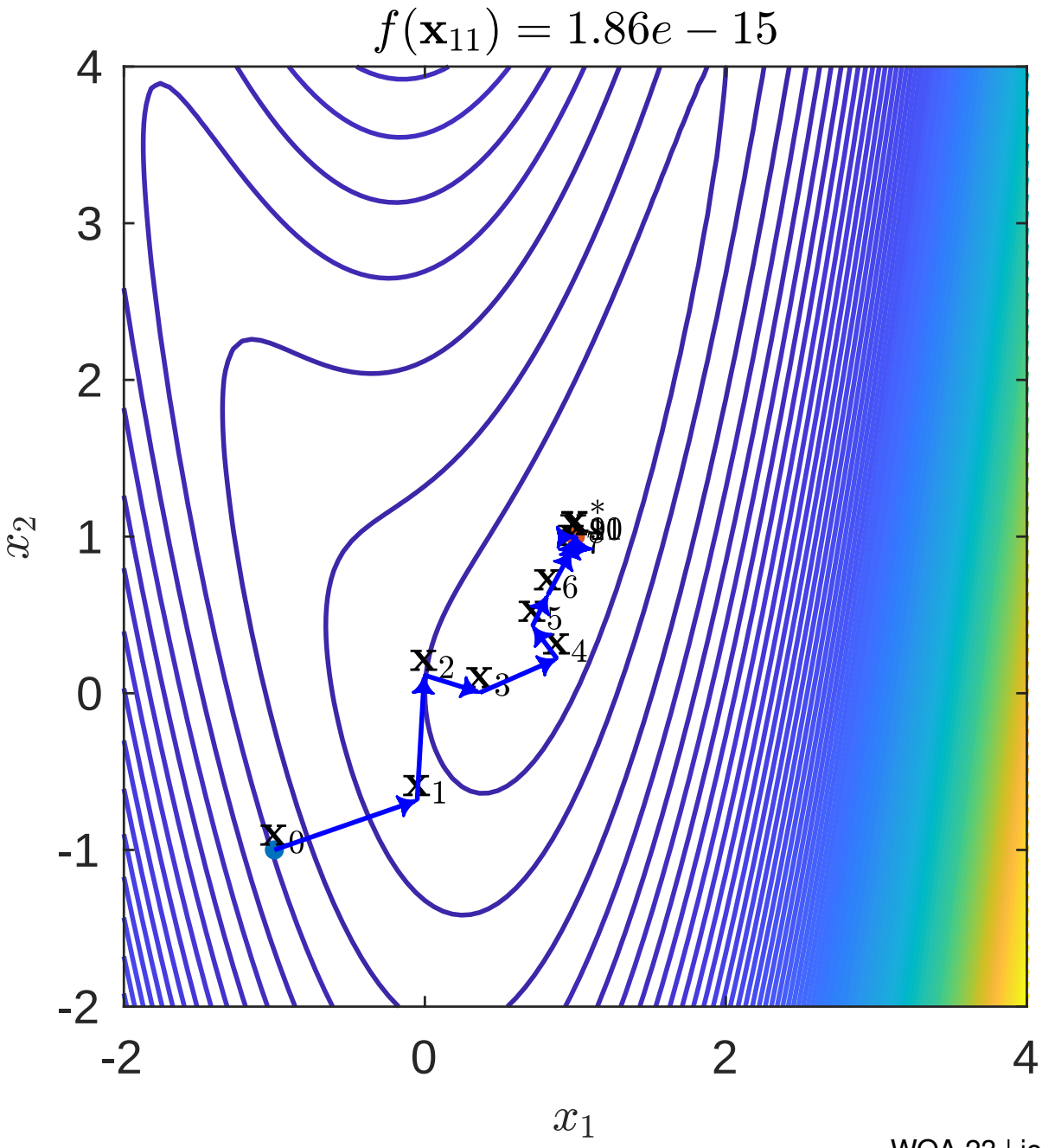
WTR-L2


$f(\mathbf{x}_{10}) = 9.28e - 10$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_{10}$
$\mathbf{x}^*$
x_1
x_2
-2
0
2
4
-1
1
3

WTR-L2


$f(\mathbf{x}_{10}) = 9.28e - 10$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_{10}$
$\mathbf{x}_{10}^*$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

WTR-L2


$f(\mathbf{x}_{11}) = 1.86e-15$
$\mathbf{x}_0$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{x}_5$
$\mathbf{x}_6$
$\mathbf{x}_{10}$
$\mathbf{x}_{11}$
$\mathbf{x}^*$
x_1
x_2
4
3
2
1
0
-1
-2
-2
0
2
4

Additional Settings

A. If $d_n/d_1 \geq 10^{16}$ restart
(conditioning)

B. Skip update if $\mathbf{s}_k^\top \mathbf{y}_k < 0$
If ≥ 20 consecutive skips restart
(definiteness)

C. If $\frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{\text{RST}}\|} \leq 10^{-2}$ and $\left(\frac{|\gamma_k|}{|\gamma_{\text{RST}}|} \geq 10 \text{ or } \frac{|\gamma_k|}{|\gamma_{\text{RST}}|} \leq 10^{-1}\right)$ restart
(only for large problems, i.e., $n > 1000$)
(scaling)

D. Initialization $\gamma_k I = B_0$ on restart $\gamma_k = \frac{\|\mathbf{g}_k\|}{\alpha_k}$
α_k is the step size of a line search after restart

Additional Settings

Count near:

If $\Delta_k \leq 10^{-4} \times \epsilon$ and

$$\|\mathbf{g}_k\|_2 \leq 5 \times 10^{-8} \|\mathbf{g}_0\|_2 \qquad \text{or}$$
$$|f_k| \leq 5 \times 10^{-11} |f_0| \qquad \text{or}$$
$$\|\mathbf{g}_k\|_2 \leq \sqrt{n} \times \epsilon$$

Parameters:

maxiter = 15000, $\epsilon = 10^{-4}$, optimal if $\|\mathbf{g}_k\|_2 < \epsilon$

Additional Settings

Count near:

If $\Delta_k \leq 10^{-4} \times \epsilon$ and

$$\|\mathbf{g}_k\|_2 \leq 5 \times 10^{-8} \|\mathbf{g}_0\|_2 \qquad \text{or}$$
$$|f_k| \leq 5 \times 10^{-11} |f_0| \qquad \text{or}$$
$$\|\mathbf{g}_k\|_2 \leq \sqrt{n} \times \epsilon$$

Parameters:

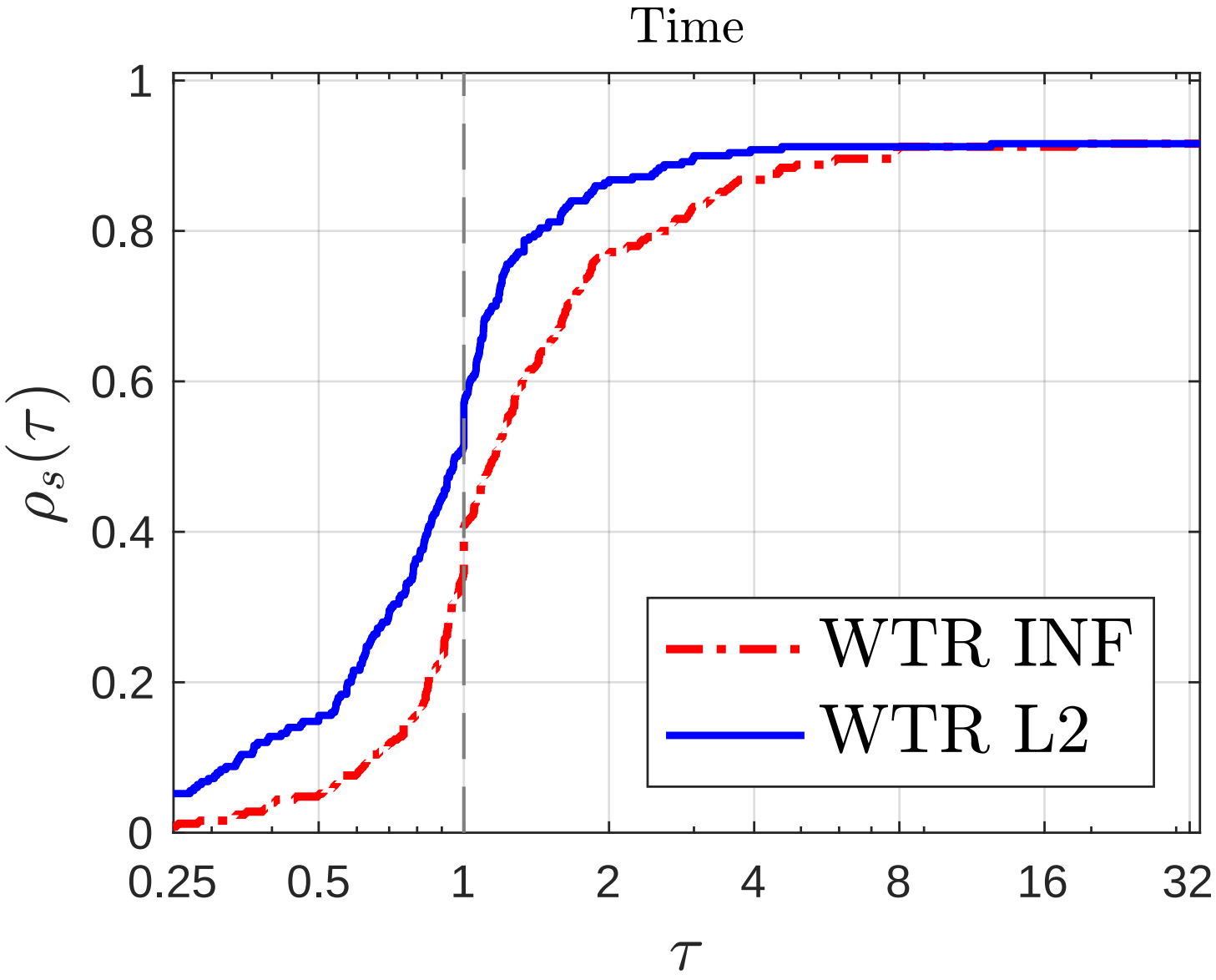
maxiter = 15000, $\epsilon = 10^{-4}$, optimal if $\|\mathbf{g}_k\|_2 < \epsilon$

Experiments on 250 CUTEst problems

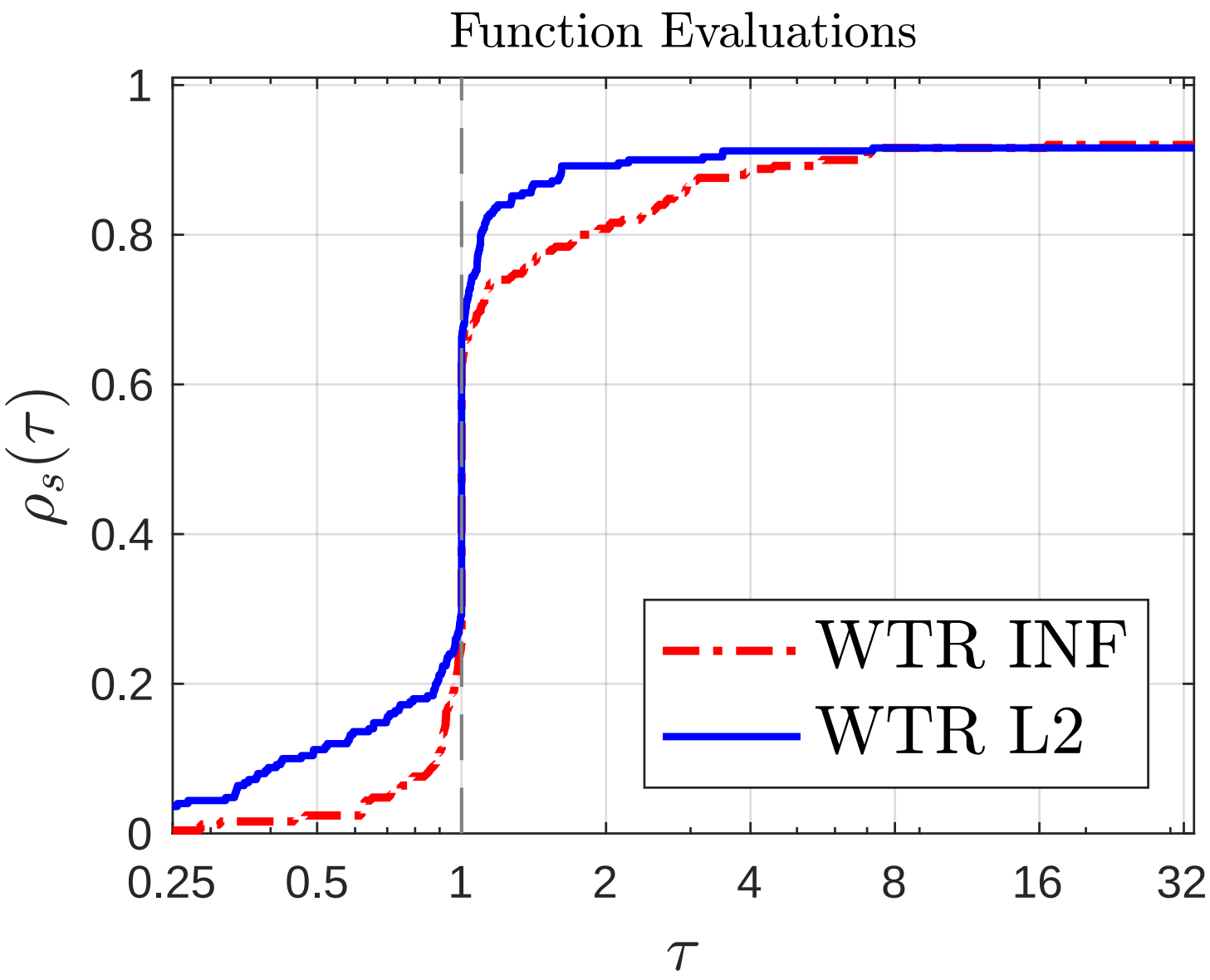
WTR-L2: First 29 CUTEst Problems

Problem	n	Iter	nF	Time	Optimal
ALLINITU	4	11	12	0.008	Optimal
ARGLINA	200	2	4	0.026	Optimal
ARGLINB	200	33	35	0.213	Near
ARGLINC	200	24	31	0.053	Near
ARGTRIGLS	200	470	472	3.354	Optimal
ARWHEAD	5000	5	10	3.239	Optimal
BA-L1LS	57	20	22	0.038	Optimal
BA-L1SPLS	57	19	21	0.046	Optimal
BDQRTIC	5000	129	134	57.147	Optimal
BEALE	2	14	15	0.007	Optimal
BENNETT5LS	3	35	37	0.023	Optimal
BIGGS6	6	41	43	0.018	Optimal
BOX	5000	28	31	12.677	Optimal
BOX3	3	9	10	0.013	Optimal
BOXBODLS	2	71	88	0.054	Optimal
BOXPOWER	5000	43	48	19.138	Optimal
BRKMCC	2	5	7	0.009	Optimal
BROWNAL	200	27	28	0.177	Optimal
BROWNBS	2	37	42	0.021	Optimal
BROWNDEN	4	42	43	0.018	Optimal
BROYDN3DLS	5000	24	66	10.646	Optimal
BROYDN7D	5000	442	449	190.664	Optimal
BROYDNBDLS	5000	56	64	24.661	Optimal
BRYBND	5000	56	64	24.778	Optimal
CERI651ALS	7	449	453	0.288	Optimal
CERI651BLS	7	325	329	0.344	Near
CERI651CLS	7	205	209	0.097	Optimal
CERI651DLS	7	280	298	0.177	Near
CERI651ELS	7	297	301	0.160	Near

250 CUTEst Problems



250 CUTEst Problems



Conclusions

- Symmetric indefinite factorization of the Hessian approximation
- Weighted L-2 and L-INF norms
- Effective subproblem solutions in the weighted norms
- Method can be robust for large class of problems

Future extension can be a trust-region line-search combination

References

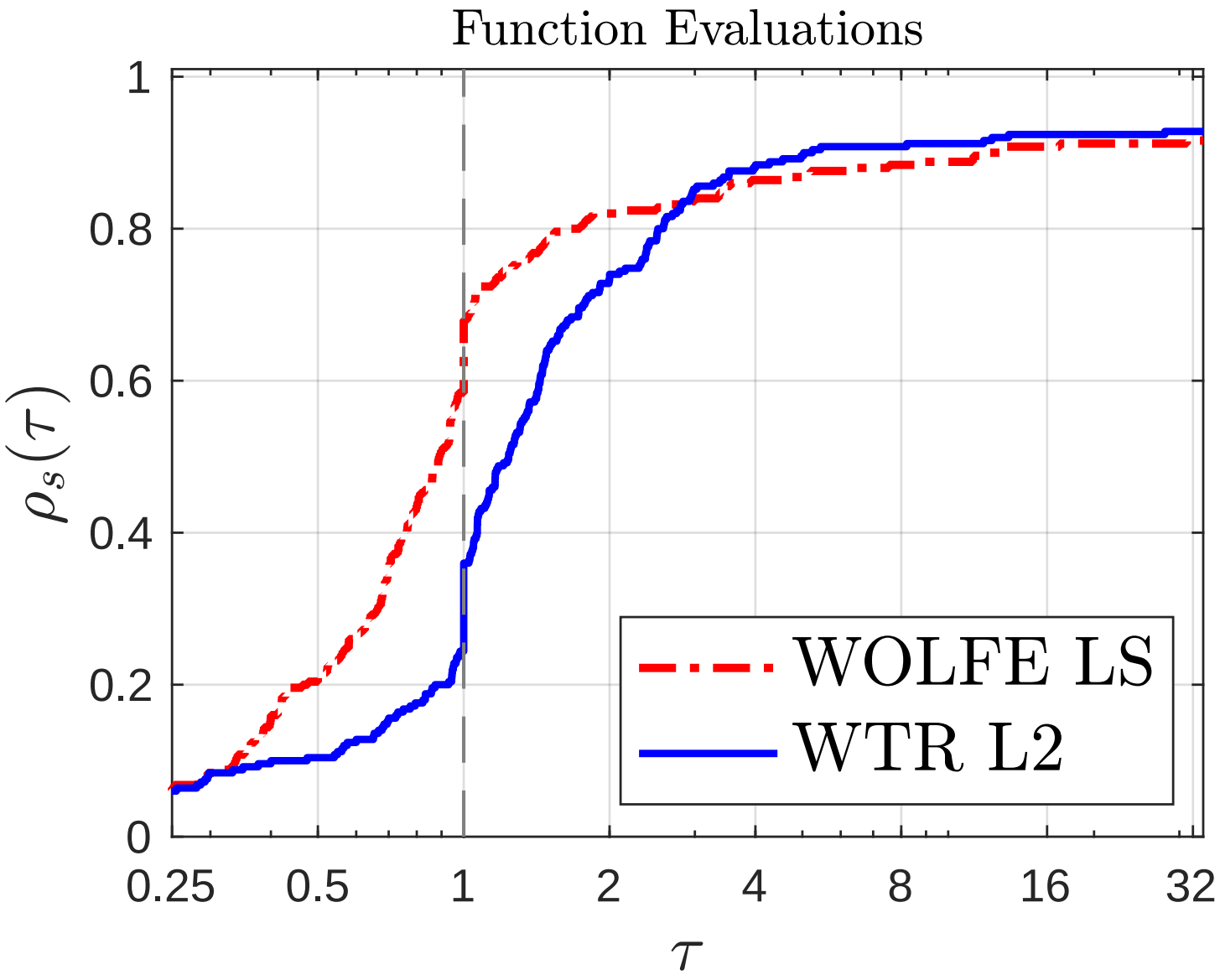
- Philip E. Gill, Gene H. Golub, Walter Murray, and Michael A. Saunders, Methods for modifying matrix factorizations, *Math. Comput.*, 28:505–535 (1974)
- Jorge J. Moré and Danny C. Sorensen, Computing a trust region step, *SIAM J. Sci. Statist. Comput.*, 4:553–572 (1983)
- Trond Steihaug, The Conjugate gradient method and trust regions in large scale optimization, *SIAM J. Numer. Anal.*, 20:626–637 (1983)
- Michael E. Gertz, A quasi-Newton trust-region method, *Mathematical Programming A*, 100:447–470 (2004)
- Jorge Nocedal and Stephen J. Wright Numerical Optimization, Springer-Verlag, New York (2006)

References

- Philip E. Gill, Gene H. Golub, Walter Murray, and Michael A. Saunders, Methods for modifying matrix factorizations, *Math. Comput.*, 28:505–535 (1974)
- Jorge J. Moré and Danny C. Sorensen, Computing a trust region step, *SIAM J. Sci. Statist. Comput.*, 4:553–572 (1983)
- Trond Steihaug, The Conjugate gradient method and trust regions in large scale optimization, *SIAM J. Numer. Anal.*, 20:626–637 (1983)
- Michael E. Gertz, A quasi-Newton trust-region method, *Mathematical Programming A*, 100:447–470 (2004)
- Jorge Nocedal and Stephen J. Wright Numerical Optimization, Springer-Verlag, New York (2006)

Thank you

Extra: Line-search comparison



Extra: Line-search comparison