The Local Convergence Properties of Pattern Search

ELIZABETH D. DOLAN
Argonne National Laboratory

ROBERT MICHAEL LEWIS
College of William & Mary

VIRGINIA TORCZON
College of William & Mary

Problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

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

Features that suggest pattern search methods are appropriate

- f is nonlinear,
- f is continuously differentiable,
- ∇f is unavailable, and
- approximations to ∇f are not reliable.

What do we ask of nonlinear optimization methods?

- Global convergence to a stationary point.
- Local rate of convergence to an isolated minimizer.
- Practical stopping criteria related to the first-order necessary conditions.

Where do we stand with respect to pattern search?

OLD Global convergence to a stationary point.

⇒ Results: Berman (1969), Céa (1971), Polak (1971), Torczon (1989), Torczon (1997), and Lewis/Torczon (1996).

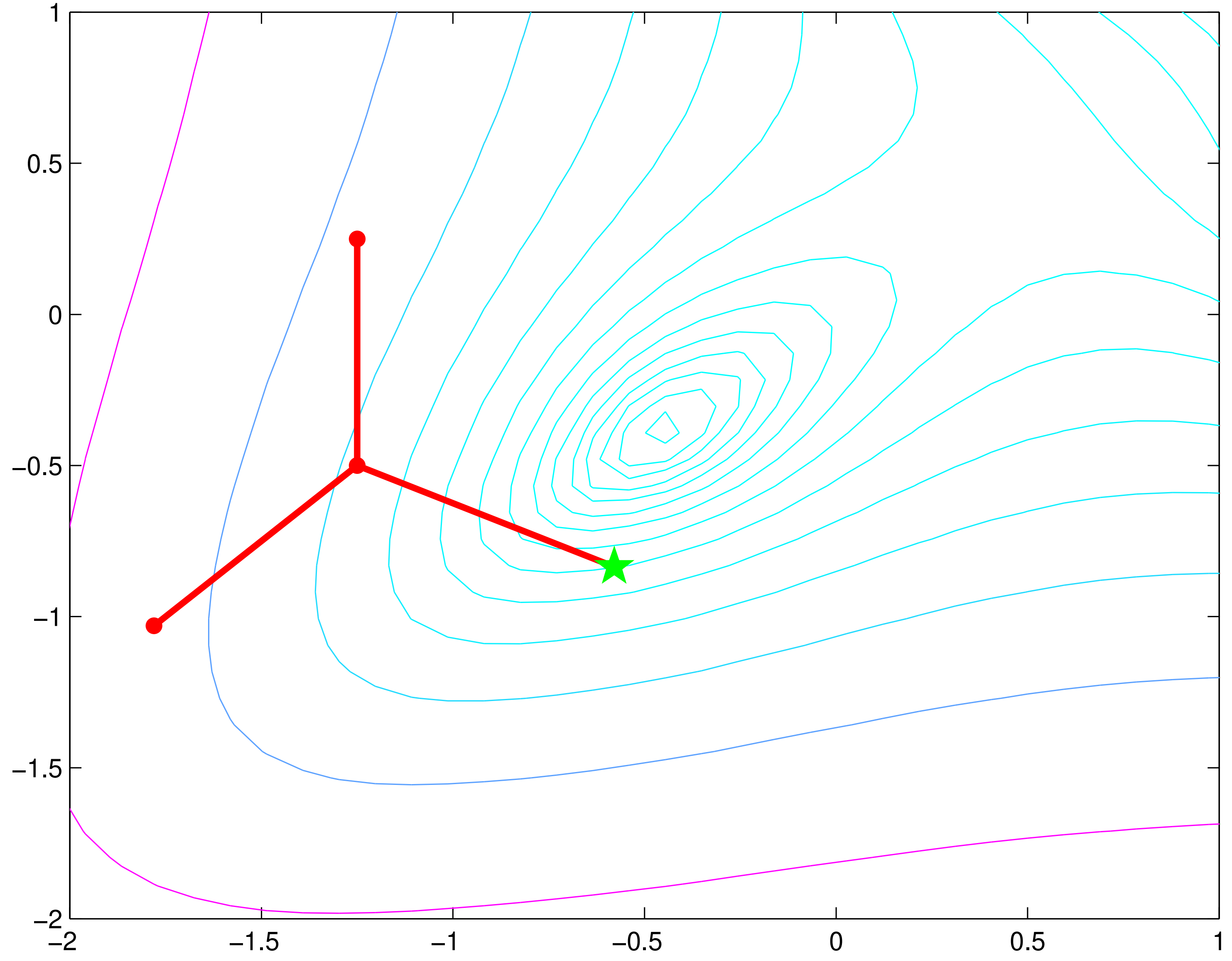
⇒ Review: Lewis, Torczon, and Trosset, *Direct search methods: then and now*, Journal of Computational and Applied Mathematics, Volume 124, Issue 1-2, December 2000, pp. 191-207.

NEW Local rate of convergence to an isolated minimizer.

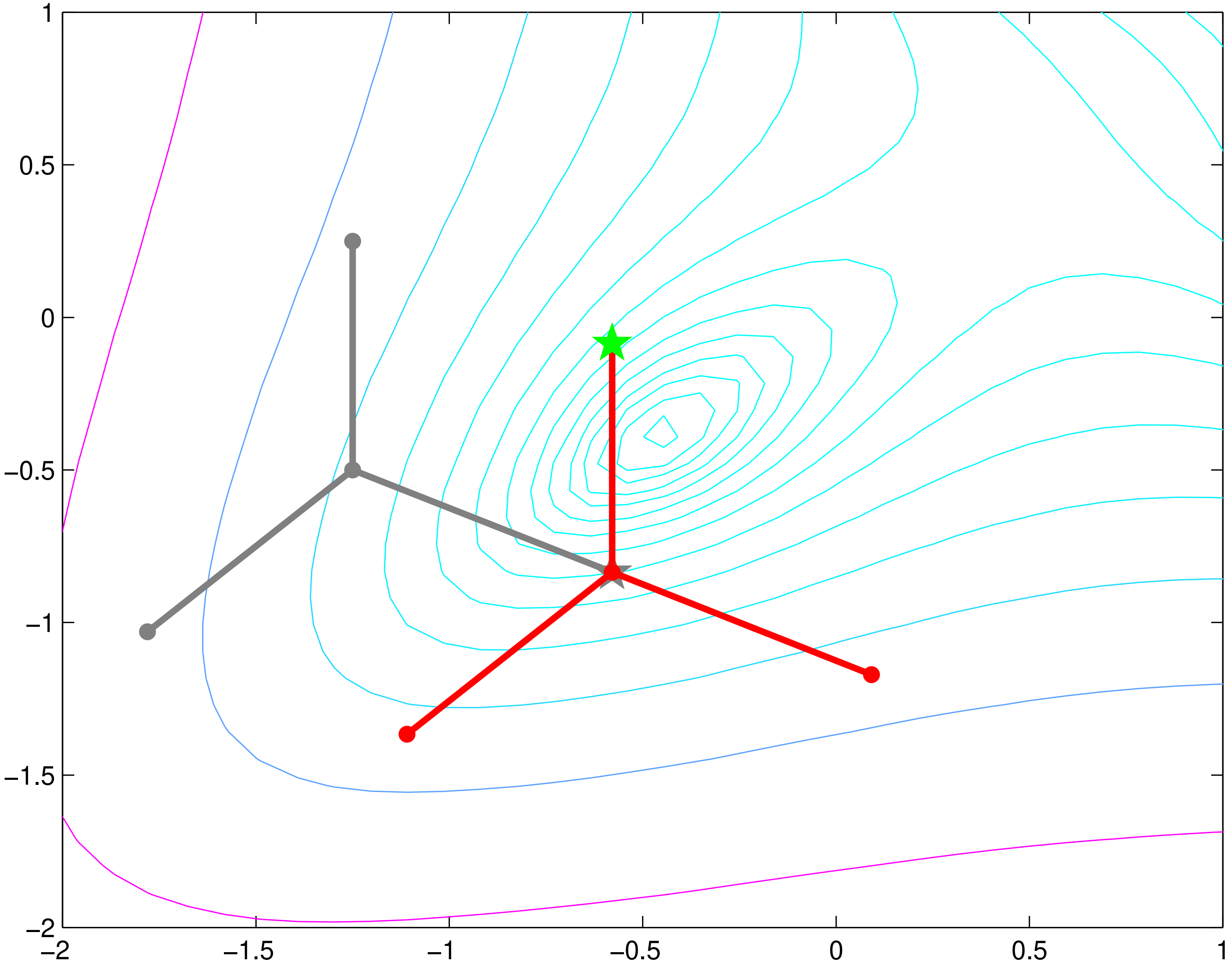
NEW Practical stopping criteria related to the first-order necessary conditions.

⇒ Report: Dolan, Lewis, and Torczon, *On the local convergence of pattern search*, ICASE Technical Report 2000-36, September 2000. In revision for SIAM Journal on Optimization.

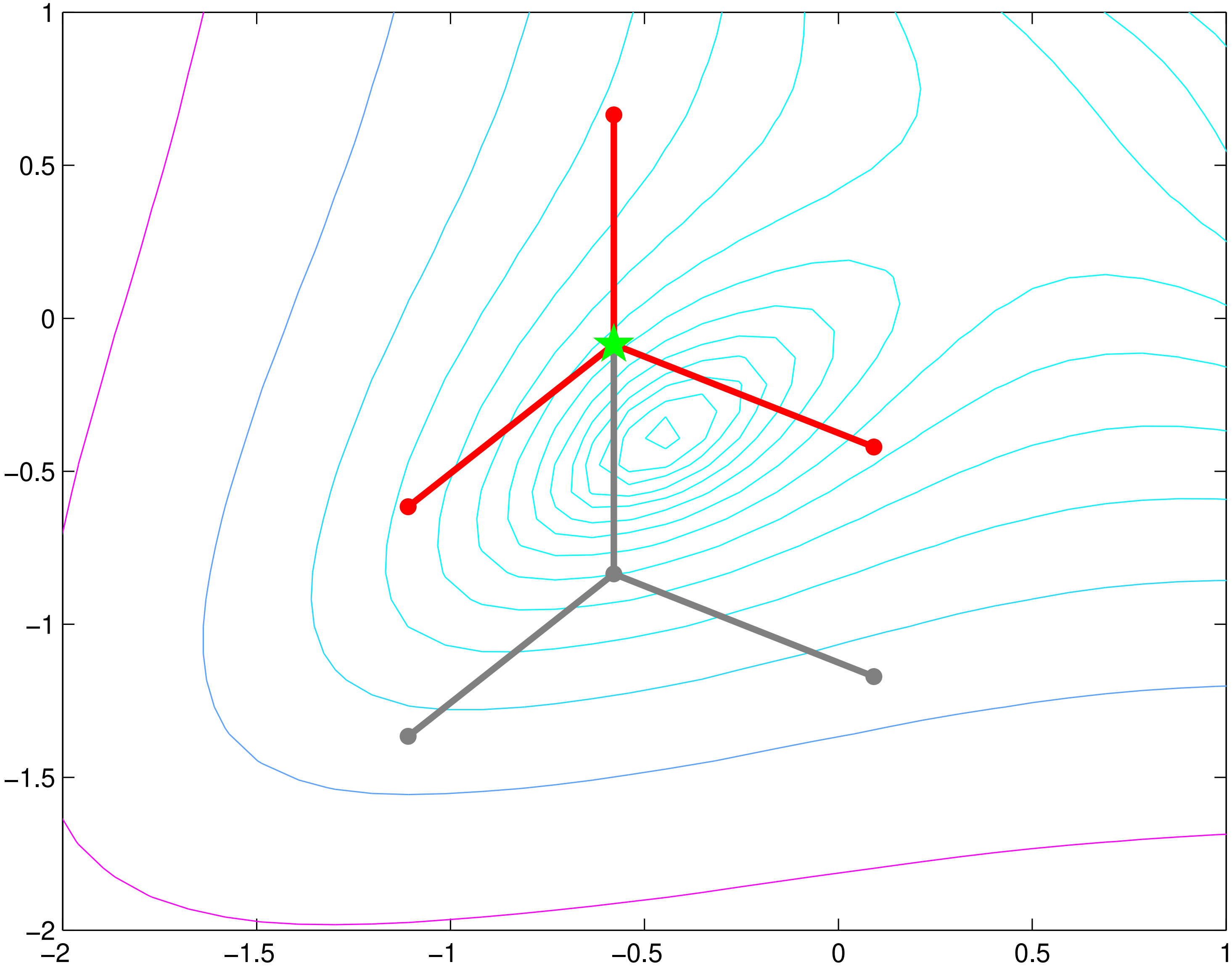
Pattern Search: first iteration is successful



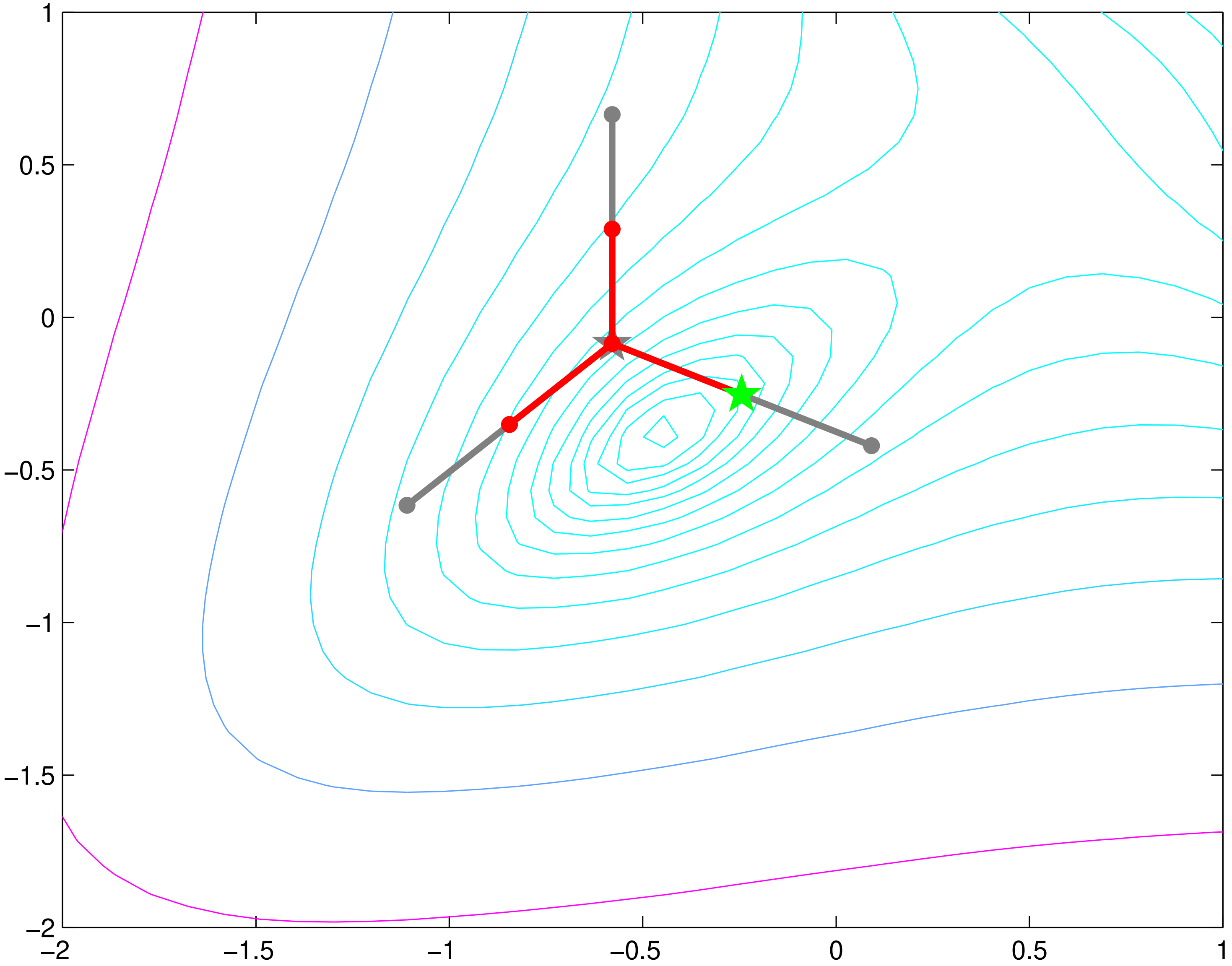
Pattern Search: second iteration is successful



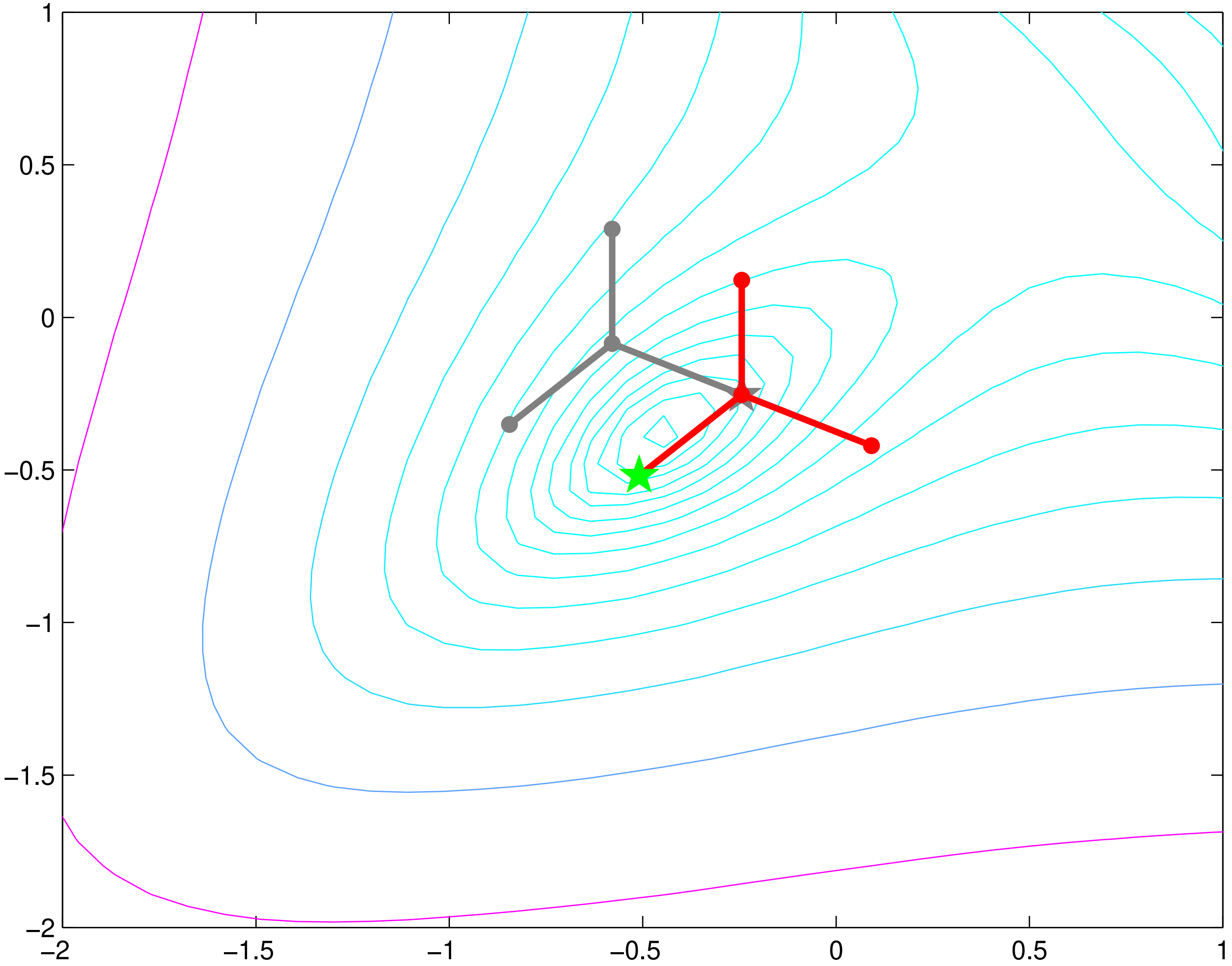
Pattern Search: third iteration is unsuccessful



Pattern Search: fourth iteration is successful



Pattern Search: fifth iteration is successful



GENERALIZED PATTERN SEARCH

Given $x_0 \in \mathbb{R}^n$, $f(x_0)$, $P_0 \in \mathbb{R}^{n \times p_0}$, $p_0 \geq n+1$, $\Delta_0 > 0$, and Δ_*, for $k = 0, 1, \ldots$ until $\Delta_k < \Delta_*$ do {

1. Find a step $s_k = \Delta_k p_k$ using the procedure `Exploratory_Moves`(Δ_k, P_k).
2. If $f(x_k + s_k) < f(x_k)$
   then $x_{k+1} = x_k + s_k$ [SUCCESSFUL ITERATION];
   otherwise, $x_{k+1} = x_k$ [UNSUCCESSFUL ITERATION].
3. `Update`(Δ_k, P_k).

}

Important observations

1. At an UNSUCCESSFUL ITERATION, where we are required to set $x_{k+1} \leftarrow x_k$ and $\Delta_{k+1} \leftarrow \frac{1}{2}\Delta_k$, we *also* know something about the *local* behavior of f; i.e.,

$$f(x_k) \quad \leq \quad f(x_k + s) \quad \forall s \in \Delta_k B\Gamma_k.$$

2. $B\Gamma_k$ is a positive spanning set.

Why are unsuccessful iterates interesting?

We can relate the inequalities

$$f(x_k) \quad \le \quad f(x_k + s) \quad \forall s \in \Delta_k B\Gamma_k.$$

to a statement about $\|\nabla f(x_k)\|$ and $\|s\|$.

In turn, $\|s\|$ is related to Δ_k.

Let's use this observation!

Focus our analysis on the subsequence of UNSUCCESSFUL ITERATES; i.e., those k's for which

- $x_{k+1} \leftarrow x_k$ and
- $\Delta_{k+1} \leftarrow \frac{1}{2}\Delta_k$.

Then we can derive two useful theorems.

Theorem 1

Given a pattern search algorithm that satisfies suitable hypotheses, suppose there exists a limit point x_* of the sequence of iterates that is a local minimizer that satisfies 2nd-order sufficiency.

Then there exist $c > 0$ and K such that for all $k \geq K$

$$\|x_k - x_*\| \leq c\Delta_{m(k)},$$

where $m(k)$ is the last unsuccessful iterate preceding k.

As a consequence, we have $\lim_{k\to\infty} x_k = x_*$.

Import of Theorem 1

For the *subsequence* of UNSUCCESSFUL ITERATES, the rate of convergence is r-linear.

THEOREM 2

Suppose $\nabla f(x)$ is Lipschitz continuous with Lipschitz constant C. Then there exist $\delta > 0$ and $c > 0$ for which the following holds.

If x_k is an iterate at which there is an unsuccessful iteration and $\Delta_k < \delta$ then

$$\|\nabla f(x_k)\| \leq c\Delta_k.$$

Import of Theorem 2

A traditional stopping criterion for pattern search —stop the search when Δ_k is "small enough"— now has sound theoretical justification.

Theorem 2 says that when Δ_k is "small enough" (i.e., "close" to 0) then $\|\nabla f(x_k)\|$ also is "small enough" (i.e., "close" to 0).

$\Rightarrow$ first-order stationarity.

Comments

- Theorem 2 is the more important of the two theorems.
- Theorem 2 says that Δ_k gives us a practical measure of $\|\nabla f(x_k)\|$ at the unsuccessful iterates.
- This measure is critical to derive Theorem 1.
- Similar results relating first-order stationarity and the size of Δ_k also hold for bound and linearly-constrained pattern search.
- This relationship is critical to derive the global convergence results for the augmented Lagrangian pattern search algorithm for problems with general constraints and simple bounds.

Further comment

Theorem 2 allows us to sharpen our earlier global convergence results.

Specifically, our earlier analysis says that, under the appropriate hypotheses,

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

If we let the set S represent a subsequence of unsuccessful iterations for which $\lim_{k\to\infty, k\in S} \Delta_k = 0$, then Theorem 2 says that we have

$$\lim_{k\to\infty, k\in S} \|\nabla f(x_k)\| = 0.$$

Final comment

Theorem 2 reassures us that, in practice, we need not worry about convergence to non-stationary points—even though the general result $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$ leaves open the possibility that $\{\|\nabla f(x_k)\|\}$ does not converge.

If we choose some suitably small stopping tolerance Δ_* and stop the algorithm at the first UNSUCCESSFUL ITERATE for which $\Delta_k < \Delta_*$, then Theorem 2 says that $\|\nabla f(x_k)\|$ also will be small.

Conclusions

For pattern search methods we now have:

— an improved global convergence result, which we see in practice;

⋆⋆ a practical measure of first-order stationarity, which we see in practice; and

— an idiosyncratic rate of convergence result ("desultory r-linear convergence"), which we see in practice.

Hooke & Jeeves Vindicated

Hooke & Jeeves (1961) had all the right intuitions regarding:

— a sufficient set of search directions,

— a sufficient mechanism for step length control, and

⋆⋆ a suitable stopping criterion.

This, despite their

> ... preference, based on experience, for straightforward search strategies which employ no techniques of classical analysis except where there is a demonstrable advantage in doing so.

To obtain the related reports

http://www.cs.wm.edu/~va
http://www-unix.mcs.anl.gov/~dolan/

In Particular

- *On the local convergence of pattern search*, Elizabeth D. Dolan, Robert Michael Lewis, and Virginia Torczon, ICASE Technical Report 2000-36, September 2000. In revision for SIAM Journal on Optimization.
- *Pattern Search Behavior in Nonlinear Optimization*, Elizabeth D. Dolan. Honors Thesis, Department of Computer Science, College of William & Mary, 1999.