Black-Box Optimisation Techniques for Complex Systems

Lindon Roberts, University of Sydney (lindon.roberts@sydney.edu.au)

Biarri Applied Mathematics Conference
29 November 2022

Outline

1. **Black-Box Optimisation**
2. Overview of optimisation techniques
3. Example application (imaging)
4. Advice

Black-Box Optimisation

A general optimisation problem looks like:

$$\min_{x} f(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where f is a single-output objective function and $\mathcal{C}$ is a feasible region.

Black-Box Optimisation

A general optimisation problem looks like:

$$\min_x f(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where f is a single-output objective function and $\mathcal{C}$ is a feasible region.

The feasible region is often described by constraint functions: $0 \leq x_i \leq 1$, $x_2 \sin(x_1) = 3$, $\sum_i x_i^2 \leq 10$, ...

Black-Box Optimisation

A general optimisation problem looks like:

$$\min_{x} f(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where f is a single-output objective function and $\mathcal{C}$ is a feasible region.

The feasible region is often described by constraint functions: $0 \leq x_i \leq 1$, $x_2 \sin(x_1) = 3$, $\sum_i x_i^2 \leq 10$, …

The decision variable(s) x might be a number, a vector or even a function of some type. There might be extra constraints implied here, e.g. integer values only.

Black-Box Optimisation

A general optimisation problem looks like:

$$\min_{x} f(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where f is a single-output objective function and $\mathcal{C}$ is a feasible region.

The feasible region is often described by constraint functions: $0 \leq x_i \leq 1$, $x_2 \sin(x_1) = 3$, $\sum_i x_i^2 \leq 10$, ...

The decision variable(s) x might be a number, a vector or even a function of some type. There might be extra constraints implied here, e.g. integer values only.

Ubiquitous in academic research & industrial problems: maximise revenue, minimise risk, maximise design efficiency, minimise prediction errors, ...

What is a black-box?

"A black-box is any proess that when provided an input, returns and output, but the inner workings of the process are not analytically available" — *Audet & Hare, 2017*

What is a black-box?

"A black-box is any proess that when provided an input, returns and output, but the inner workings of the process are not analytically available" — Audet & Hare, 2017

Black-box functions often arise when:

- Function evaluation involves real-world activities (e.g. perform a lab experiment)
- Legacy/proprietary code is involved in a computation
- Any sufficiently complex calculation/simulation can effectively be a black-box ("I don't have time to figure this out!")

Black-Box Optimisation

What is a black-box?

"A black-box is any proess that when provided an input, returns and output, but the inner workings of the process are not analytically available" — *Audet & Hare, 2017*

Black-box functions often arise when:

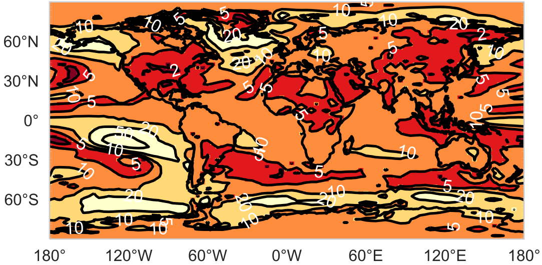
- Function evaluation involves real-world activities (e.g. perform a lab experiment)
- Legacy/proprietary code is involved in a computation
- Any sufficiently complex calculation/simulation can effectively be a black-box ("I don't have time to figure this out!")

Black-box optimisation = optimisation when at least one objective/constraint function is a black box.

Application 1: Climate Modelling

[Tett et al., 2022]

- Parameter calibration for global climate models (least squares minimisation)
- One model run = simulate global climate for 5 years (expensive!)
- Very complicated, chaotic physics (black-box & noisy!)



Application 2: Adversarial Example Generation [Alzantot et al., 2019]

- Find perturbations of neural network inputs which are misclassified (min. probability of correct label/max. probability of desired incorrect label)
- Neural network structure assumed to be unknown (black-box!)
- Want to test very few examples ($\approx$ expensive!)

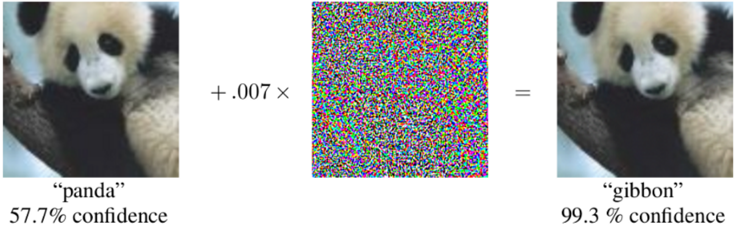


Image from [Goodfellow et al., 2015]

Black-Box Optimisation

Issue?

Why is black-box optimisation hard?

Generally two reasons black-box optimisation may be difficult:

- Evaluating the BB functions may be very slow/costly or inaccurate
- Don't have access to derivatives of BB functions (and hard to estimate)
  - Some recent work considers only *comparison oracles*: given x and y, say which is better (but no values of f), e.g. survey results [Slavin & McKenzie, 2022]

Black-Box Optimisation

Issue?

Why is black-box optimisation hard?

Generally two reasons black-box optimisation may be difficult:

- Evaluating the BB functions may be very slow/costly or inaccurate
- Don't have access to derivatives of BB functions (and hard to estimate)
  - Some recent work considers only *comparison oracles*: given x and y, say which is better (but no values of f), e.g. survey results [Slavin & McKenzie, 2022]

Most standard optimisation techniques assume availability of derivatives of objective and all constraint functions, and will evaluate functions at many points to achieve high-quality solutions.

Outline

1. Black-Box Optimisation
2. **Overview of optimisation techniques**
3. Example applications (imaging)
4. Advice

Overview of Techniques

For simplicity, focus on methods for unconstrained or box-constrained problems.

Other constrained problems usually solved with modifications of these underlying techniques.

Overview of Techniques

For simplicity, focus on methods for unconstrained or box-constrained problems.

Other constrained problems usually solved with modifications of these underlying techniques.

Important distinction: local or global optimisation?

- Local: best x amongst nearby points
  - Faster, more theoretical guarantees.
  - Better when many decision variables.
- Global: best x in entire feasible region
  - Can find better points, but usually slower.
  - Only theory is "works if you eventually search everywhere".
  - Algorithms balance "exploration" (search new areas) with "exploitation" (zoom in on known good regions).

Model-Based Methods

Very successful framework: model-based methods

- Evaluate f at an initial collection of points
- Build an approximation to f (e.g. interpolation). Can be local or global models (region of accuracy)
- Minimise approximation to select a new point to evaluate
  - If global, sometimes consider far away points instead (exploration)
- Evaluate new point and update model

Model-Based Methods

Very successful framework: model-based methods

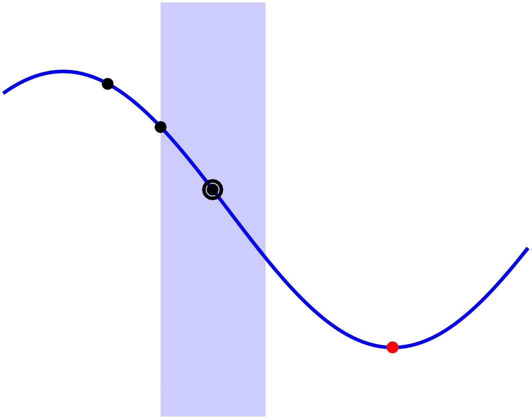
- Evaluate f at an initial collection of points
- Build an approximation to f (e.g. interpolation). Can be local or global models (region of accuracy)
- Minimise approximation to select a new point to evaluate
  - If global, sometimes consider far away points instead (exploration)
- Evaluate new point and update model

Comes in local (model-based derivative-free optimisation) and global (surrogate/Bayesian optimisation) flavours.

I have several open-source Python software packages of this type!

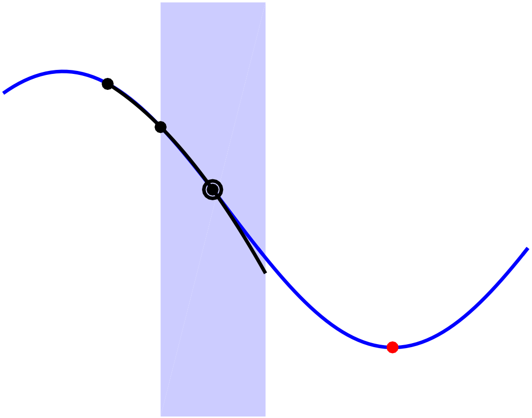
[Conn, Scheinberg & Vicente, 2009; Shahriari et al., 2016]

Example: Model-Based DFO



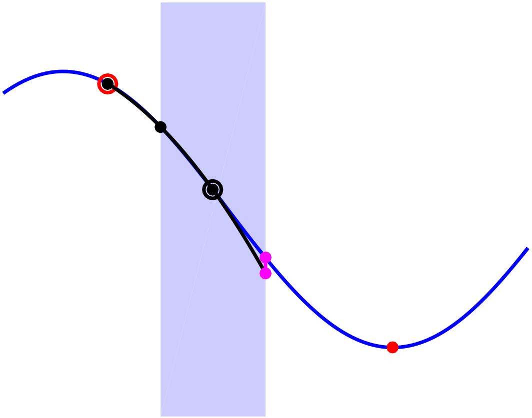
1. Choose interpolation set

Example: Model-Based DFO



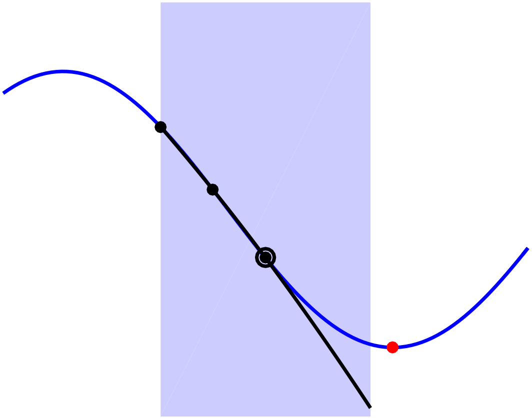
2. Interpolate & minimize...

Example: Model-Based DFO



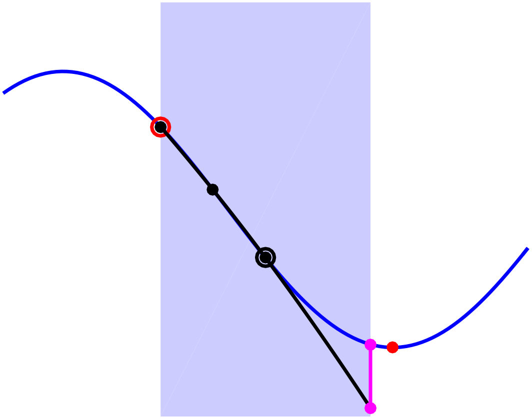
3. Add new point to interpolation set (replace a bad point)

Example: Model-Based DFO



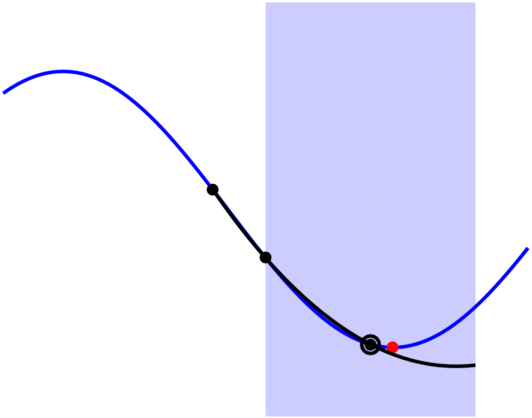
4. Repeat with new interpolation set & model

Example: Model-Based DFO



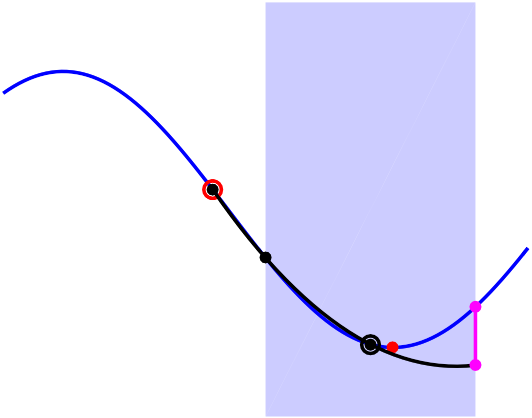
4. Repeat with new interpolation set & model

Example: Model-Based DFO



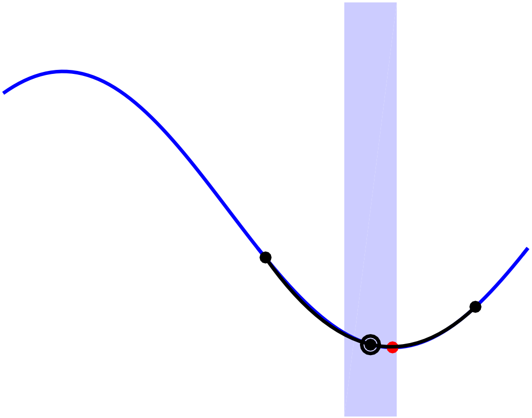
4. Repeat with new interpolation set & model

Example: Model-Based DFO



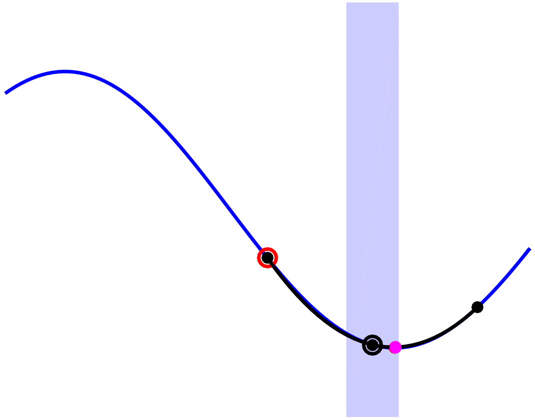
4. Repeat with new interpolation set & model

Example: Model-Based DFO



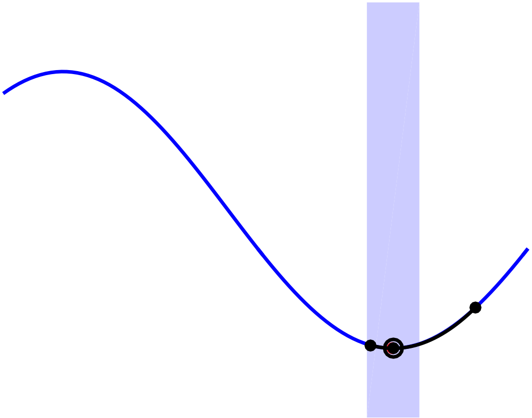
4. Repeat with new interpolation set & model

Example: Model-Based DFO



4. Repeat with new interpolation set & model

Example: Model-Based DFO



4. Repeat with new interpolation set & model

Global method: DIRECT (DIviding RECTangles)

- Start with one rectangle: the feasible box region (constraints assumed)
- Evaluate f in the centre of all rectangles
- Pick rectangle(s) which are large and/or have small centre f
- Subdivide these rectangles into thirds to get new rectangles

Rapidly gained popularity and many modifications proposed to improve performance.

[Jones, Perttunen & Stuckman, 1993]

Example: DIRECT

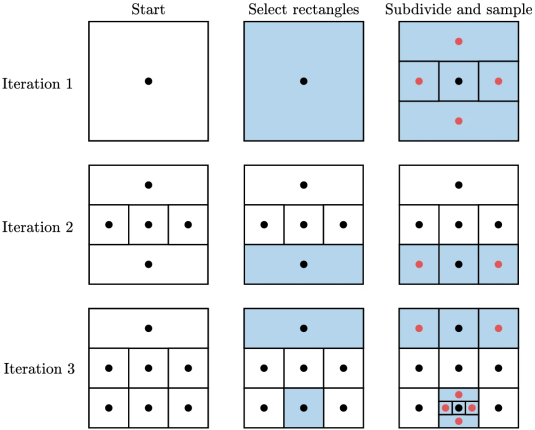


Image from [Jones & Martins, 2021]

Other Methods

Many other methods not discussed here, such as

- Nelder-Mead simplex method (local)
- Direct search (local)
- Simulated annealing (global)
- Genetic/evolutionary algorithms (global)
- Particle swarm (global)

Other Methods

Many other methods not discussed here, such as

- Nelder-Mead simplex method (local)
- Direct search (local)
- Simulated annealing (global)
- Genetic/evolutionary algorithms (global)
- Particle swarm (global)

Nelder-Mead very popular because of its simplicity, not so good in practice. Direct search mostly works well if augmented with model-based ideas, but simple/flexible.

Global methods 'inspired by nature' tend not to perform as well as 'inspired by mathematics' (but quite general, best as fallback option).

"Methods inspired by nature are for the ignorant or the desperate." — A. R. Conn, 2018.

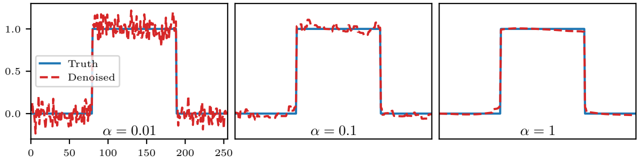
Outline

1. Black-Box Optimisation
2. Overview of optimisation techniques
3. **Example application (imaging)**
4. Advice

Inverse Problems

A common inverse problem is to match some observed data, encouraging solutions of a particular sort with a regulariser.

e.g. Image denoising: given a noisy image y, find a denoised image $\hat{x}$ by minimising $\text{error}(\hat{x}, y) + \alpha\,\text{variation}(\hat{x})$. The solution depends on the choice of α.



How to choose good problem parameters?

Learn from data! Given $(x_1, y_1), \ldots, (x_n, y_n)$ — ground truth and noisy observations, find α which minimises average $\text{error}(\hat{x}_i(\alpha), x_i)$. Treat $\hat{x}_i(\alpha)$ as black box!

Results

Can learn parameters which give good denoising results:

Takeaways: black-box optimisation realistic (don't pretend you have derivatives you can't compute), faster results than assuming derivatives.

Next step: learn sampling pattern for MRIs (i.e. which Fourier coefficients to collect).

Outline

1. Black-Box Optimisation
2. Overview of optimisation techniques
3. Example application (imaging)
4. **Advice**

General Advice

My advice when doing optimisation on complex systems/black-box functions:

- Pick your variables, objective and constraints carefully — what do you actually want to solve?
- Understand the structure: discrete vs. continuous variables, linear vs. nonlinear constraints, special problem structure (e.g. least-squares)
- Local vs. global methods?
  - Local is faster, more scalable, better guarantees, better at exploiting good starting guesses & problem structure
  - Global can sometimes find better solutions
- Software? [Rios & Sahinidis, 2013; Cartis, R. & Sheridan-Methven, 2022]
  - Local: mine! (see my website)
  - Global: DIRECT or PySOT [Regis & Shoemaker, 2007]

Conclusions

Key Message

Optimisation of complex, black-box systems is possible, but it takes a lot of work.

Conclusions

Key Message

Optimisation of complex, black-box systems is possible, but it takes a lot of work.

Conclusions

- Standard optimisation packages (SciPy, fmincon, ...) usually won't work well
- Alternative techniques exist and rapidly becoming more sophisticated
- Big ongoing research area: scalability (large numbers of decision variables, e.g. machine learning)

References i

M. Alzantot, Y. Sharma, S. Chakraborty, H. Zhang, C.-J. Hsieh, and M. B. Srivastava, *GenAttack: Practical black-box attacks with gradient-free optimization*, in Proceedings of the Genetic and Evolutionary Computation Conference, Prague, Czech Republic, 2019, ACM, pp. 1111–1119.

C. Cartis, L. Roberts, and O. Sheridan-Methven, *Escaping local minima with local derivative-free methods: A numerical investigation*, Optimization, 71 (2022), pp. 2343–2373.

A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*, vol. 8 of MPS-SIAM Series on Optimization, MPS/SIAM, Philadelphia, 2009.

I. J. Goodfellow, J. Shlens, and C. Szegedy, *Explaining and harnessing adversarial examples*, in 3rd International Conference on Learning Representations ICLR, San Diego, 2015.

D. R. Jones and J. R. R. A. Martins, *The DIRECT algorithm: 25 years later*, Journal of Global Optimization, 79 (2021), pp. 521–566.

D. R. Jones, C. D. Perttunen, and B. E. Stuckman, *Lipschitzian optimization without the Lipschitz constant*, Journal of Optimization Theory and Applications, 79 (1993), pp. 157–181.

References ii

T. G. Kolda, R. M. Lewis, and V. Torczon, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Review, 45 (2003), pp. 385–482.

R. G. Regis and C. A. Shoemaker, *A stochastic radial basis function method for the global optimization of expensive functions*, INFORMS Journal on Computing, 19 (2007), pp. 497–509.

L. M. Rios and N. V. Sahinidis, *Derivative-free optimization: A review of algorithms and comparison of software implementations*, Journal of Global Optimization, 56 (2013), pp. 1247–1293.

B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. D. Freitas, *Taking the human out of the loop: A review of Bayesian optimization*, Proceedings of the IEEE, 104 (2016), pp. 148–175.

I. Slavin and D. McKenzie, *Adapting zeroth order algorithms for comparison-based optimization*, arXiv preprint arXiv:2210.05824, (2022).

S. F. B. Tett, J. M. Gregory, N. Freychet, C. Cartis, M. J. Mineter, and L. Roberts, *Does model calibration reduce uncertainty in climate projections?*, Journal of Climate, 35 (2022), pp. 2585–2602.

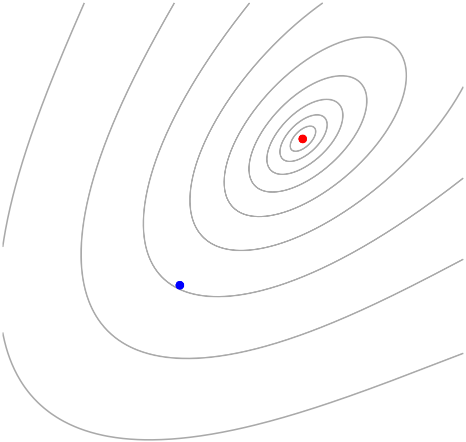
Simple local method: direct search

- Look at fixed perturbations around best point so far (for a given step length)
- If some perturbation gives sufficient improvement:
  - Make perturbation new best point, increase step length
- Otherwise, keep old best point, decrease step length

Need perturbations to satisfy: for any fixed vector v, there is always a perturbation d making an acute angle with v. For example, $\{\pm$coordinate dirctions$\}$.

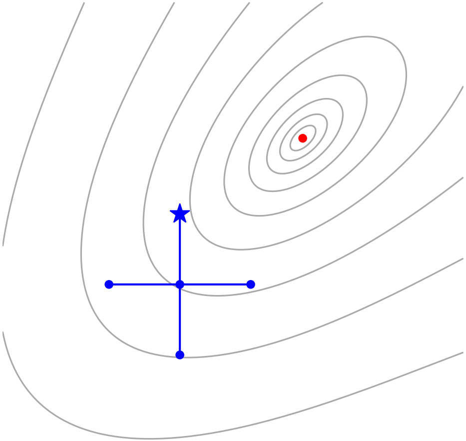
[Kolda, Lewis & Torczon, 2003; Conn, Scheinberg & Vicente, 2009]

Example: Direct Search



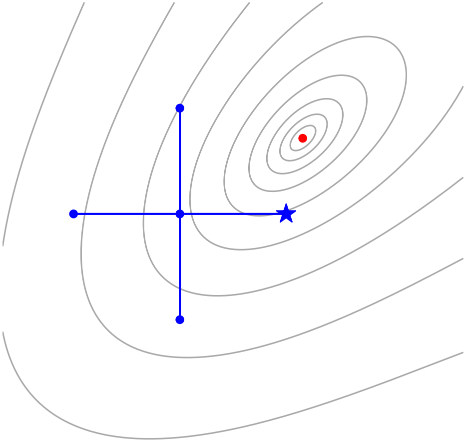
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



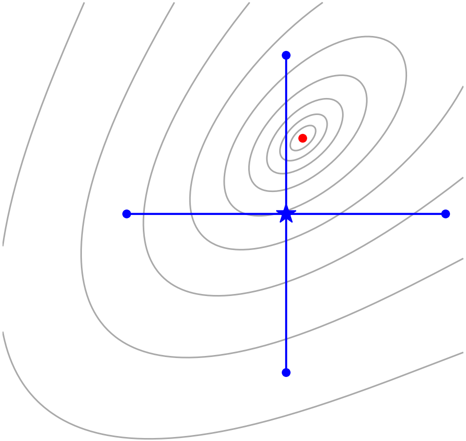
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



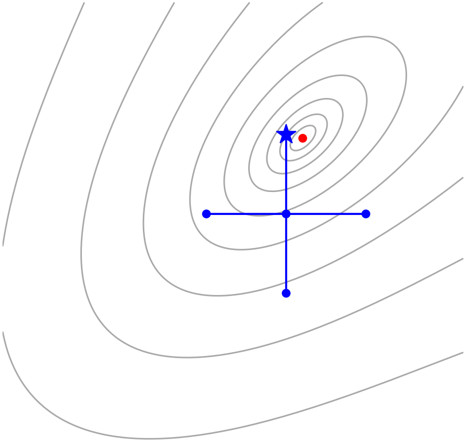
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



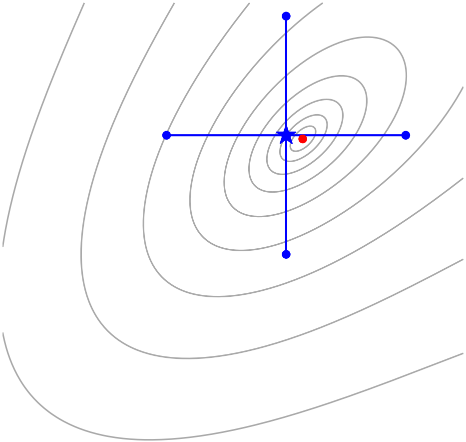
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



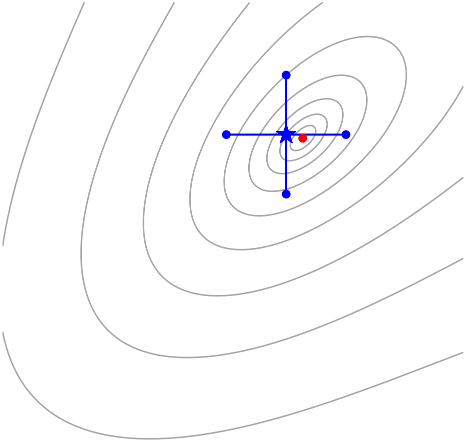
Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



Modified from [Kolda, Lewis & Torczon, 2003]

Example: Direct Search



Modified from [Kolda, Lewis & Torczon, 2003]