Generating Set Search for Nonlinear Programming

Virginia Torczon, College of William & Mary

Collaborators:

- Tammy Kolda, Sandia National Laboratory
- Michael Lewis, College of William & Mary

The goals of the talk:

- To illustrate a simple GSS method.
- To use this illustration to derive a general form for GSS methods.
- To illustrate the features of GSS methods that ensure convergence.
- To show that at an identifiable subsequence of $\{x_k\}$, there is an implicit bound on the norm of the gradient in terms of the step-length control parameter Δ_k.

  It is from this result that both the global and local convergence results follow.

What are Generating Set Search (GSS) methods?

Look at one of the simplest possible examples compass search applied to the problem:

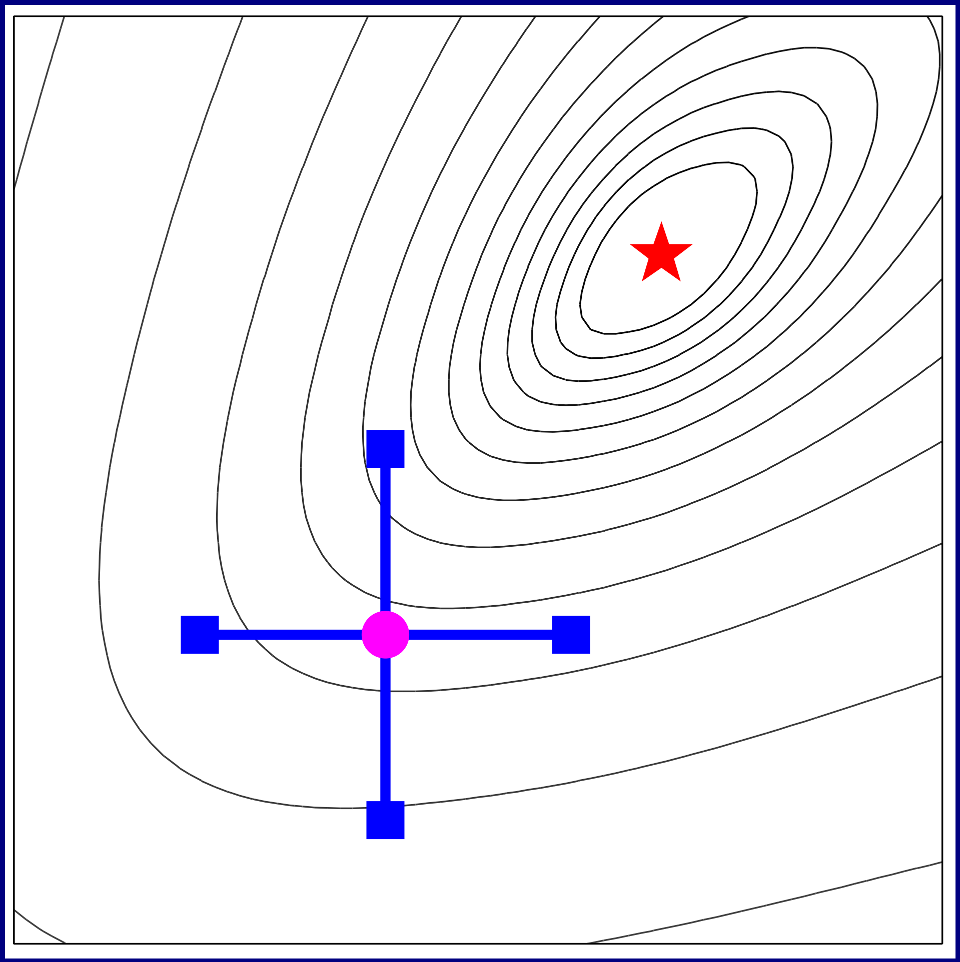
$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \;\; f(x^1, x^2)$$

where

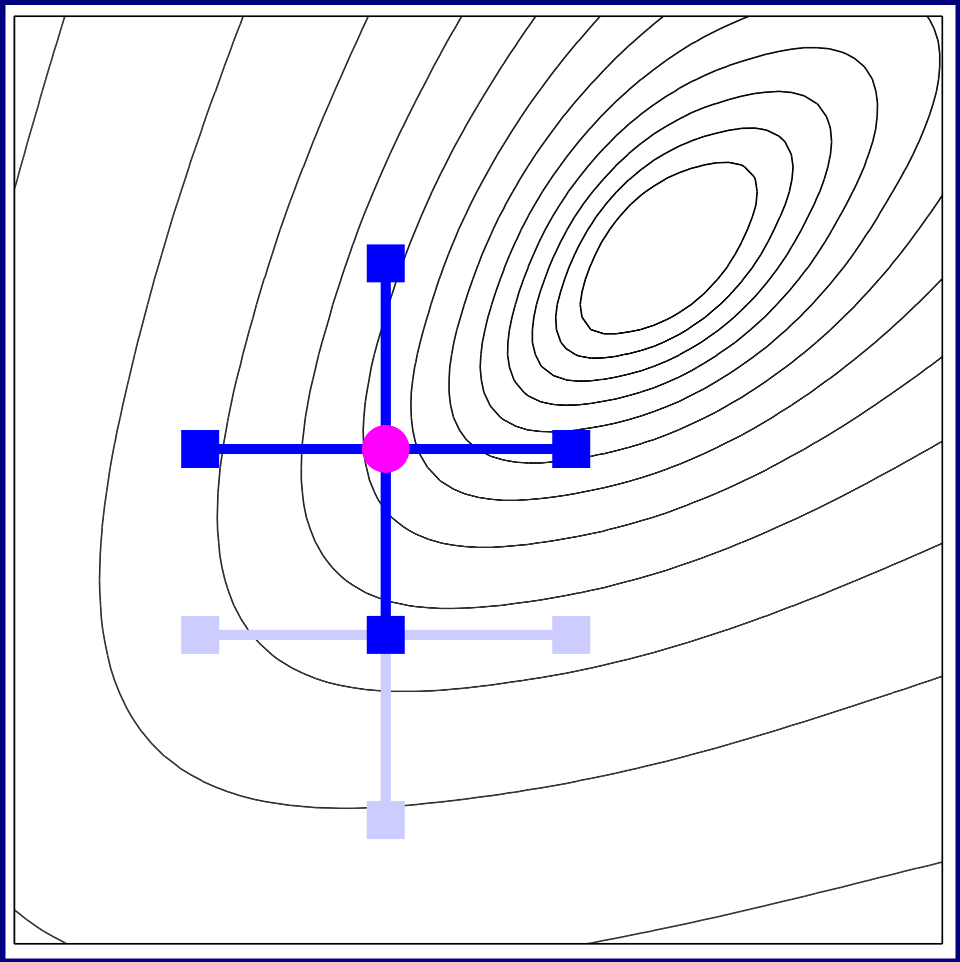
$$f(x) = \left|(3 - 2x^1)x^1 - 2x^2 + 1\right|^{\frac{7}{3}} + \left|(3 - 2x^2)x^2 - x^1 + 1\right|^{\frac{7}{3}},$$

(the modified Broyden tridiagonal function).

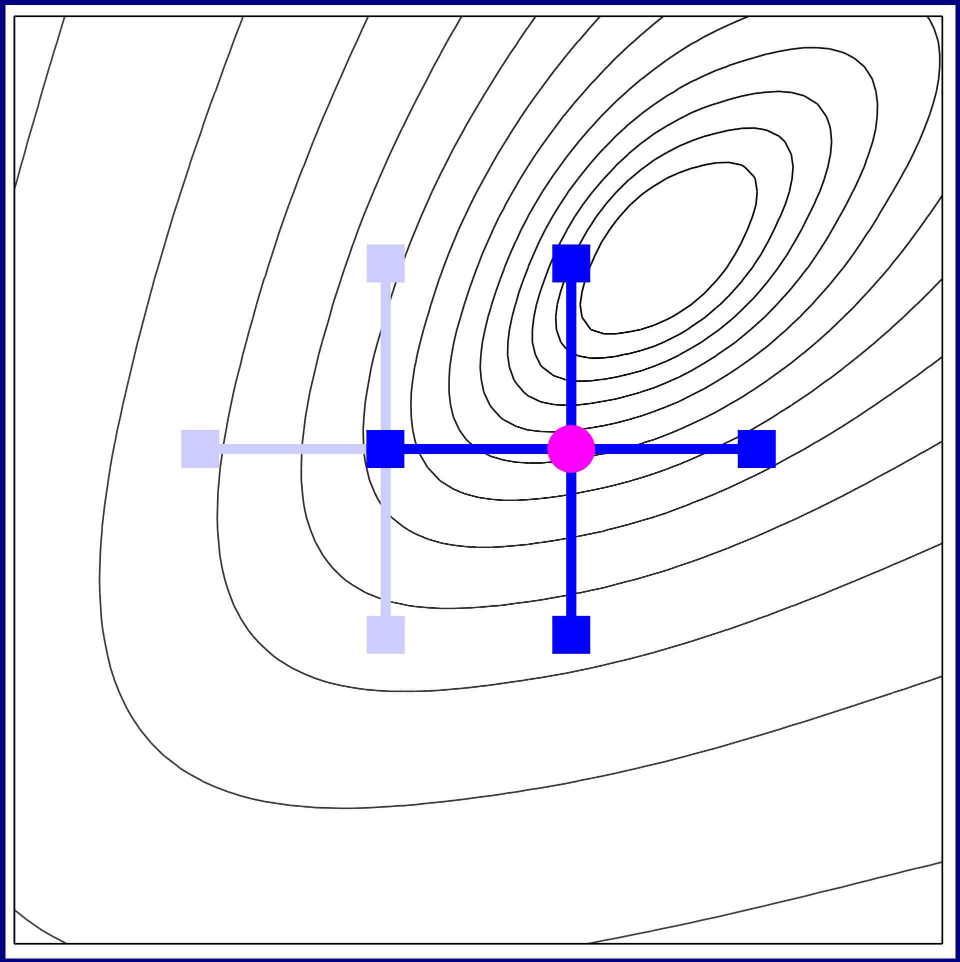
Initial pattern:



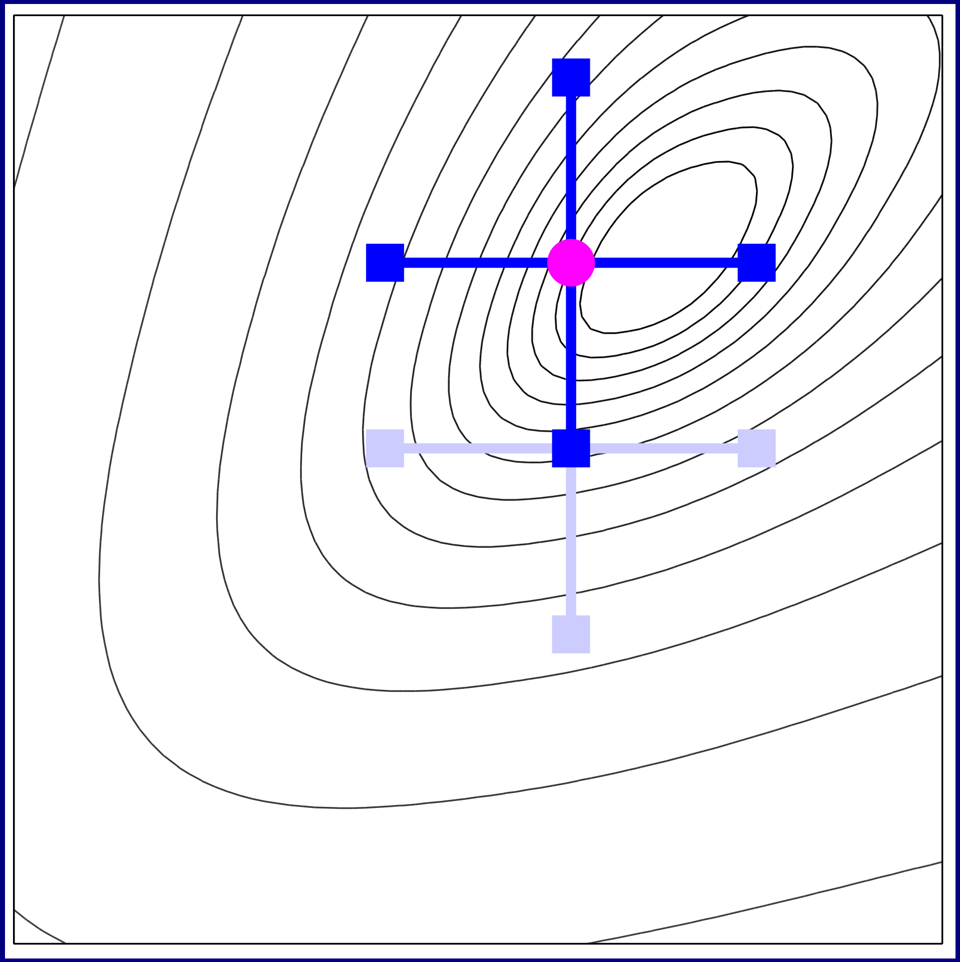
Move North:



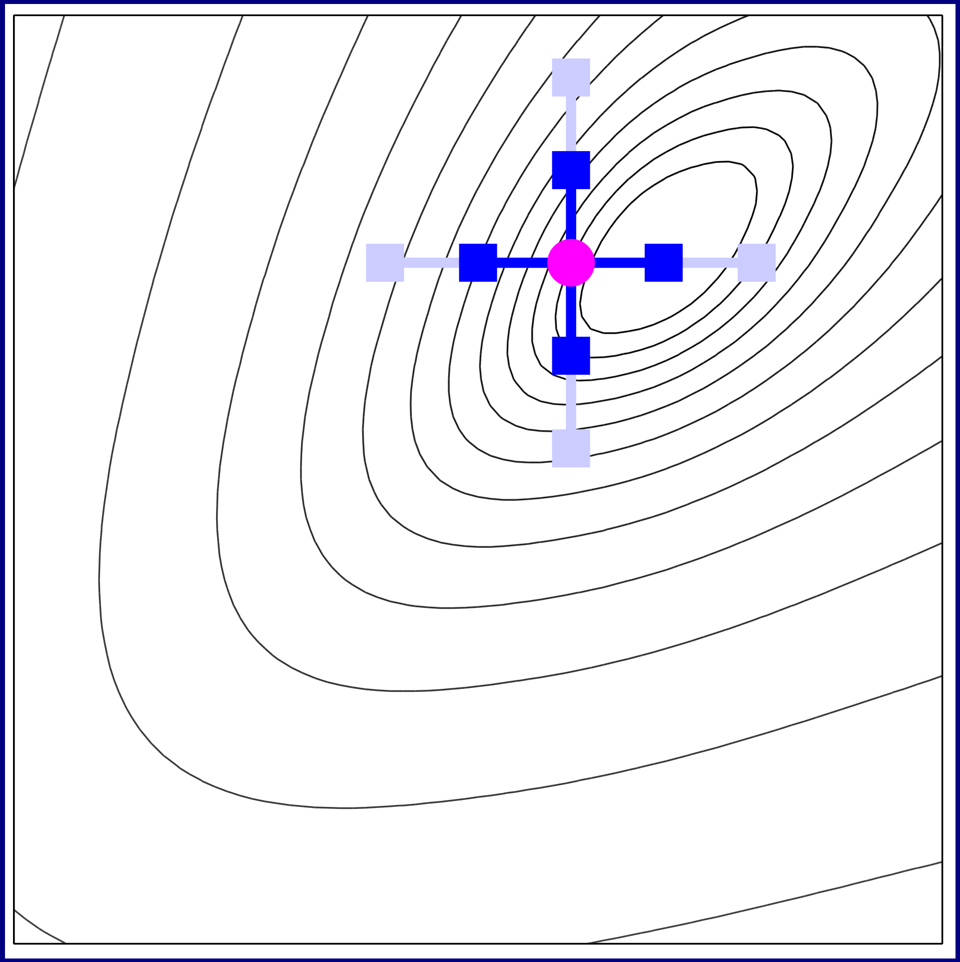
Move West:



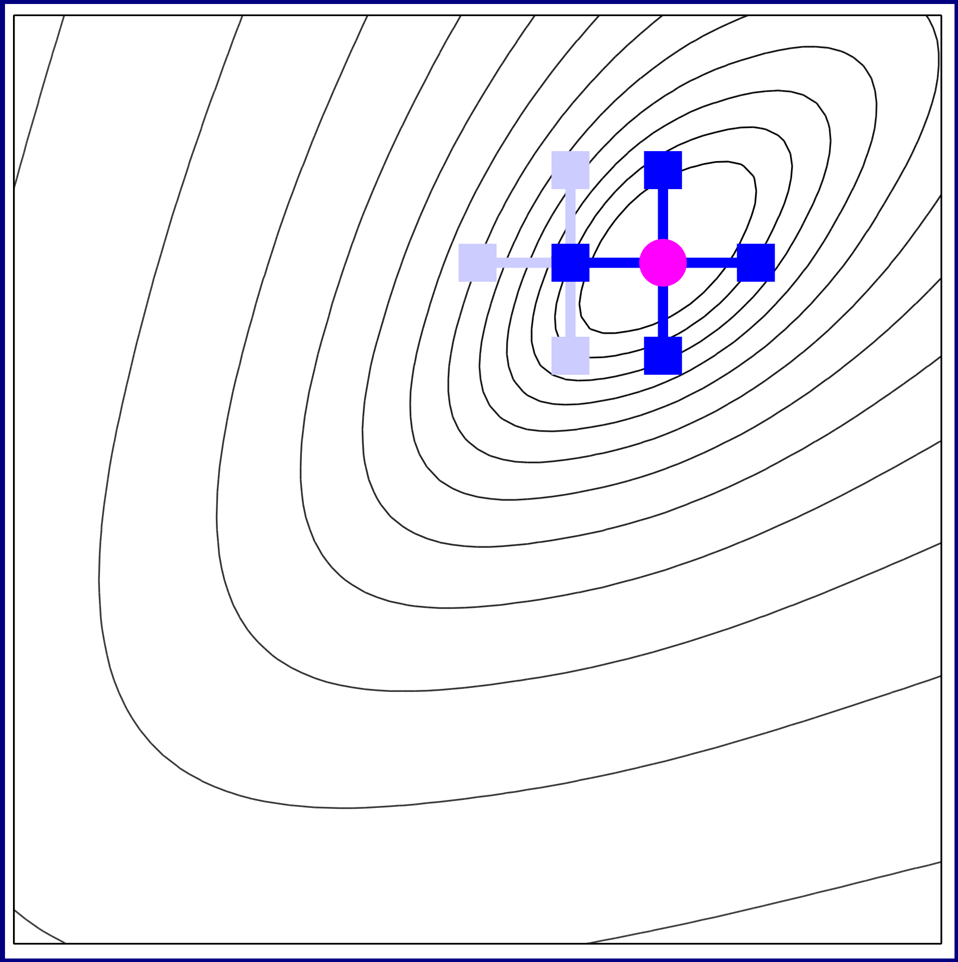
Move North:



Contract:



Move West:



Compass search: initialization

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be given.

Let $x_0 \in \mathbb{R}^n$ be the initial guess.

Let $\Delta_{\text{tol}} > 0$ be the tolerance used to test for convergence.

Let $\Delta_0 > \Delta_{\text{tol}}$ be the initial value of the step length control parameter.

Compass search: algorithm

For each iteration $k = 1, 2, \ldots$

Step 1. Let $\mathcal{D}_\oplus$ be the set of coordinate directions $\{\pm e_i \,|\, i = 1, \ldots, n\}$, where e_i is the ith unit coordinate vector in $\mathbb{R}^n$.

Step 2. If there exists $d_k \in \mathcal{D}_\oplus$ such that $f(x_k + \Delta_k d_k) < f(x_k)$ then the iteration is *successful*.

Do the following:

- Set $x_{k+1} = x_k + \Delta_k d_k$ (change the iterate).
- Set $\Delta_{k+1} = \Delta_k$ (no change to the step length control parameter).

Step 3. Otherwise, $f(x_k + \Delta_k d) \geq f(x_k)$ for all $d \in \mathcal{D}_{\oplus}$, so the iteration is *unsuccessful*.

Do the following:

- Set $x_{k+1} = x_k$ (no change to the iterate).
- Set $\Delta_{k+1} = \frac{1}{2}\Delta_k$ (contract the step length control parameter).
- If $\Delta_{k+1} < \Delta_{\text{tol}}$, then **terminate**.

What is needed to ensure global convergence?

There are two basic conditions:

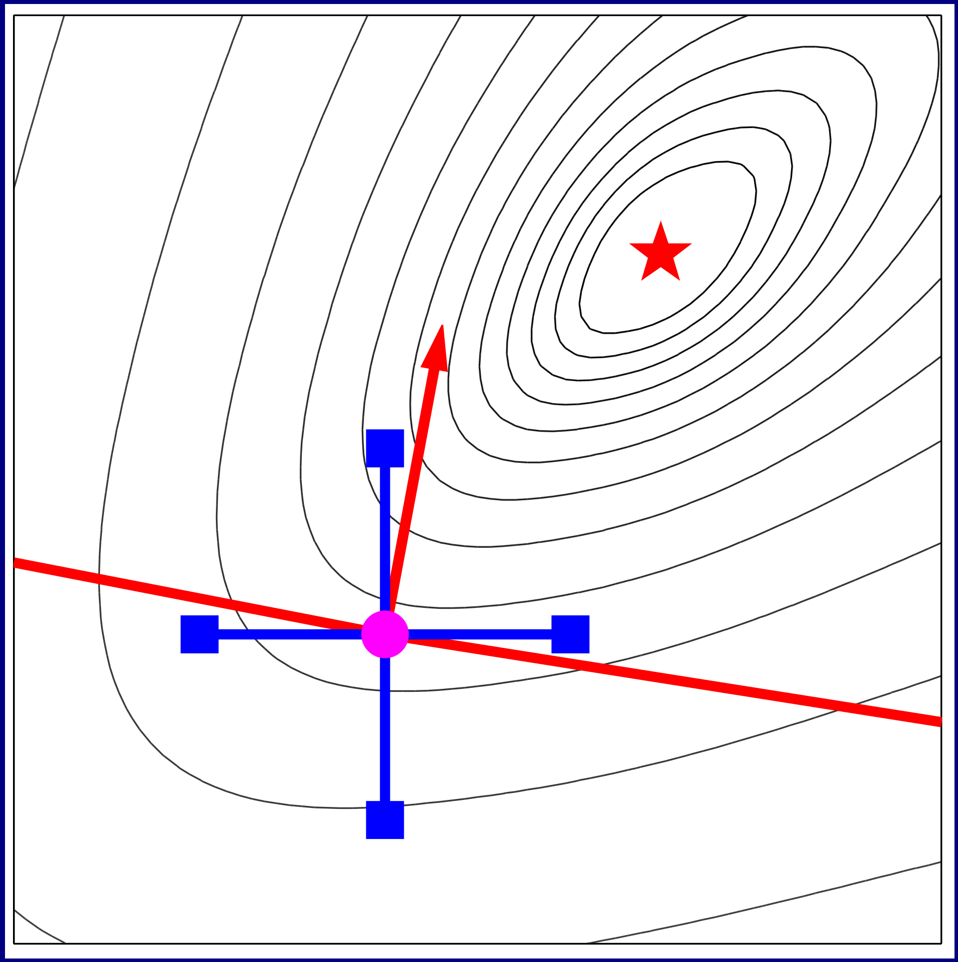
- A reasonable direction of descent.
- A reasonable choice of step length along that direction of descent to ensure that the step is neither
  - ⋆ "too long" relative to the amount of decrease seen from one iterate to the next nor
  - ⋆ "too short" relative to the linear rate of decrease in the function.

What makes GSS methods interesting analytically?

The typical safeguards for nonlinear optimization make explicit use of $\nabla f(x_k)$ to ensure a reasonable choice of search direction and step length.

If we assume ∇f exists and is continuous, it is possible to construct GSS methods that satisfy these conditions without explicitly using $\nabla f(x)$.

Compass search guarantees a direction of descent:



Specifically, compass search guarantees:

The cosine of the largest angle between an arbitrary vector $v \in \mathbb{R}^n$ and the closest coordinate direction in the set $\mathcal{D}_\oplus$ is bounded below by $\frac{1}{\sqrt{n}}$.

Thus, no matter the value of $\nabla f(x)$, there is at least one $d \in \mathcal{D}_\oplus$ for which

$$\kappa(\mathcal{D}_\oplus) = \frac{1}{\sqrt{n}} \leq \frac{-\nabla f(x_k)^T d}{\| \nabla f(x_k) \| \, \| d \|}.$$

Extending the observation to GSS:

We can replace the set of coordinate direction $\mathcal{D}_\oplus$ with a set of search directions $\mathcal{D}_k$. The conditions on $\mathcal{D}_k$ are

- that $\mathcal{D}_k$ contain a *generating set* for $\mathbb{R}^n$ and
- that $\mathcal{D}_k$ satisfies an angle condition of the form $\kappa(\mathcal{D}_k) \geq \kappa_{\min} > 0$.

Generating sets for $\mathbb{R}^n$

Let $\mathcal{G}$ denote a set of p directions in $\mathbb{R}^n$, with the ith direction denoted by d^i. Then we say that $\mathcal{G}$ *generates* (or *positively spans*) $\mathbb{R}^n$ if for any vector $v \in \mathbb{R}^n$, there exist $\lambda^1, \ldots, \lambda^p \geq 0$ such that

$$v = \sum_{i=1}^{p} \lambda^i \, d^i.$$

Clearly the set of coordinate directions:

$$\mathcal{D}_{\oplus} = \{\pm e_i \,|\, i = 1, \ldots, n\}$$

satisfies this condition.

But there is an infinite number of other algorithmic possibilities.

The generating set guarantees a direction of descent

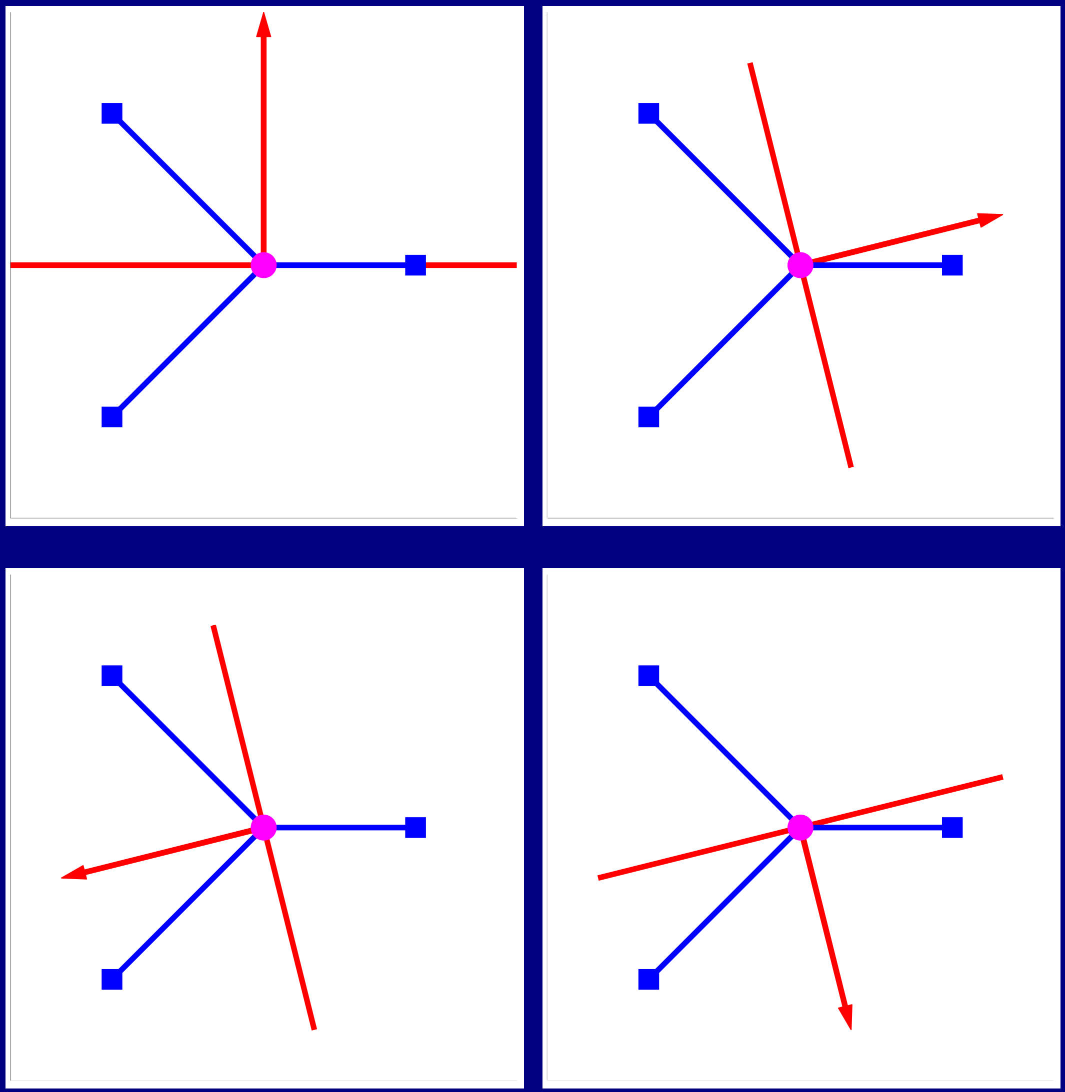
Lemma. *The set $\mathcal{G}$ generates $\mathbb{R}^n$ if and only if for any vector $v \in \mathbb{R}^n$ such that $v \neq 0$, there exists $d \in \mathcal{G}$ such that $v^T d > 0$.*

Geometrically, this says the $\mathcal{G}$ generates $\mathbb{R}^n$ if and only if the interior of every half-space contains a member of $\mathcal{G}$.

The significance to GSS is that if at every iteration k, $\mathcal{D}_k$ contains a generating set for $\mathbb{R}^n$, then there must be at least one $d \in \mathcal{D}_k$ such that

$$-\nabla f(x_k)^T d > 0.$$

Thus, $\mathcal{D}_k$ contains at least one direction of descent whenever $\nabla f(x_k) \neq 0$.

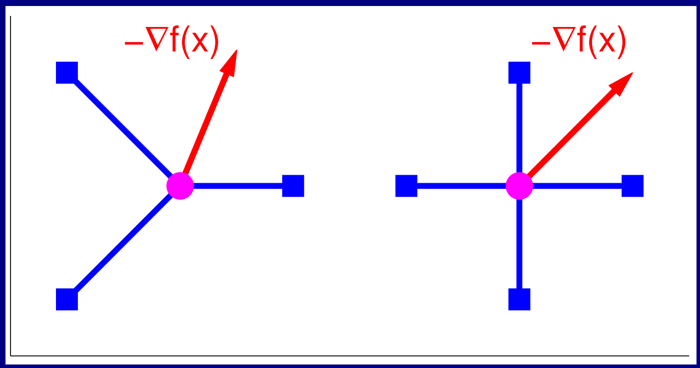


A measure of the quality of the direction of descent

Formally, the *cosine measure* of $\mathcal{G}$ is:

$$\kappa(\mathcal{G}) \equiv \min_{v \in \mathbb{R}^n} \max_{d \in \mathcal{G}} \frac{v^T d}{\|v\| \, \|d\|}.$$

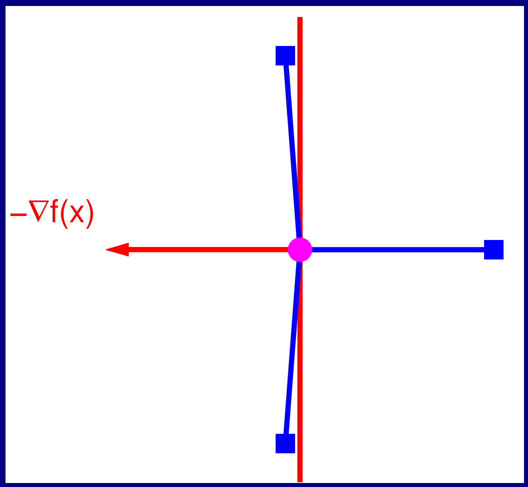
This measure captures how far the steepest descent direction can be, in the worst case, from the vector d in $\mathcal{G}$ making the smallest angle with $v = -\nabla f(x)$.



$\kappa(\mathcal{G})$ is required to be uniformly bounded below

$$\kappa(\mathcal{G}_k) \geq \kappa_{\min} > 0 \quad \text{for all} \quad k = 1, 2, \ldots.$$

This lower bound is meant to prevent pathologies such as



thus ensuring a reasonable direction of descent.

How to ensure a reasonable choice of step length?

Use a step-length control parameter Δ_k and for a given $d_k \in \mathcal{D}_k$, only accept the step $\Delta_k d_k$ if

$$f(x_k + \Delta_k d_k) < f(x_k) - \rho(\Delta_k),$$

where $\rho : \mathbb{R} \to \mathbb{R}$ is a nonnegative function such that $\rho(t)/t \to 0$ as $t \to 0$.

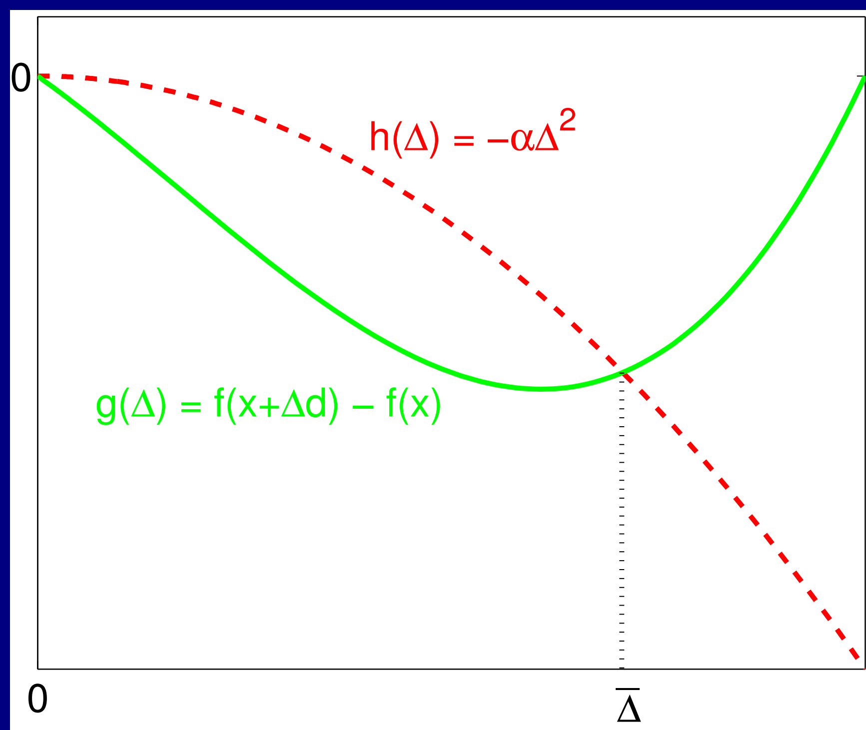
Two choices:

- $\rho \equiv 0$ (simple decrease)
- $\rho(t) = \alpha t^2$ for some $\alpha > 0$ (sufficient decrease)

Why $\rho(t)/t \to 0$ as $t \to 0$?

For *success* require:

$$f(x_k + \Delta_k d_k) - f(x_k) < -\rho(\Delta_k).$$



GSS: initialization

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be given.

Let $x_0 \in \mathbb{R}^n$ be the initial guess.

Let $\Delta_{\mathrm{tol}} > 0$ be the step length convergence tolerance.

Let $\Delta_0 > \Delta_{\mathrm{tol}}$ be the initial value of the step length control parameter.

Let $\theta_{\mathrm{max}} < 1$ be an upper bound on the contraction parameter.

Let $\rho : \mathbb{R} \rightarrow \mathbb{R}$ be a nonnegative function such that $\rho(t)/t \rightarrow 0$ as $t \rightarrow 0$. The choice $\rho \equiv 0$ is acceptable.

Let β_{min} and β_{max} be lower and upper bounds, respectively, on the lengths of the vectors in any generating set.

Let κ_{min} be a lower bound on the cosine measure of any generating set.

GSS: algorithm

For each iteration $k = 1, 2, \ldots$

Step 1. Let $\mathcal{D}_k = \mathcal{G}_k \cup \mathcal{H}_k$. Here $\mathcal{G}_k$ is a generating set for $\mathbb{R}^n$ satisfying $\beta_{\min} \leq \|d\| \leq \beta_{\max}$ for all $d \in \mathcal{G}_k$ and $\kappa(\mathcal{D}_k) \geq \kappa_{\min}$, and $\mathcal{H}_k$ is a finite (possibly empty) set of additional search directions such that $\beta_{\min} \leq \|d\|$ for all $d \in \mathcal{H}_k$.

Step 2. If there exists $d_k \in \mathcal{D}_k$ such that $f(x_k + \Delta_k d_k) < f(x_k) - \rho(\Delta_k)$, then the iteration is *successful*.

Do the following:

- Set $x_{k+1} = x_k + \Delta_k d_k$ (change the iterate).
- Set $\Delta_{k+1} = \phi_k \Delta_k$, where $\phi_k \geq 1$ (optionally expand the step length control parameter).

Step 3. Otherwise, $f(x_k + \Delta_k d) \geq f(x_k) - \rho(\Delta_k)$ for all $d \in \mathcal{D}_k$, so the iteration is *unsuccessful*.

Do the following:

- Set $x_{k+1} = x_k$ (no change to the iterate).
- Set $\Delta_{k+1} = \theta_k \Delta_k$ where $0 < \theta_k < \theta_{\max} < 1$ (contract the step length control parameter).
- If $\Delta_{k+1} < \Delta_{\text{tol}}$, then **terminate**.

Relating Δ_k to the measure of stationarity

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, and suppose ∇f is Lipschitz continuous with constant M. Then GSS produces iterates such that for any $k \in \mathcal{U}$, we have*

$$\| \nabla f(x_k) \| \leq \frac{1}{\kappa(\mathcal{G}_k)} \left[M \Delta_k \beta_{\max} + \frac{\rho(\Delta_k)}{\Delta_k \beta_{\min}} \right].$$

For simplicity we assume ∇f is Lipschitz but this can be relaxed to the assumption that ∇f is only continuously differentiable.

Globalization

We can ensure that, at the very least, GSS methods produce iterations satisfying

$$\liminf_{k \to +\infty} \Delta_k = 0.$$

At least three mechanisms:

- Globalization via a rational lattice ($\rho = 0$) [Torczon]
- Globalization via moving grids ($\rho = 0$) [Coope/Price]
- Globalization via a sufficient decrease condition ($\rho \neq 0$) [Lucidi/Sciandrone]

One consequence:

If we require

$$\rho(t)/t \longrightarrow 0 \text{ as } t \longrightarrow 0$$

and we can show

$$\liminf_{k \to +\infty} \Delta_k = 0,$$

then

$$\| \nabla f(x_k) \| \leq \frac{1}{\kappa(\mathcal{G}_k)} \left[M \Delta_k \beta_{\max} + \frac{\rho(\Delta_k)}{\Delta_k \beta_{\min}} \right].$$

ensures that, at the very least,

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

BOTTOM LINE: GSS methods are *globally convergent*.

Other consequences:

In addition, the stationarity measure, together with the choice of an appropriate globalization strategy and some stronger assumptions, leads to local convergence results:

- $\lim_{k \to +\infty} x_k = x_*$.
- For an identifiable subsequence of $\{x_k\}$, $\| x_k - x_* \| \leq c\Delta_k$ for some c independent of k.
- This identifiable subsequence of $\{x_k\}$ is r-linearly convergent.

Furthermore:

The relationship between Δ_k and $\| \nabla f(x_k) \|$:

$$\| \nabla f(x_k) \| \leq \frac{1}{\kappa(\mathcal{G}_k)} \left[M \Delta_k \beta_{\max} + \frac{\rho(\Delta_k)}{\Delta_k \beta_{\min}} \right]$$

means that Δ_k is an appropriate stopping criterion to test after an *unsuccessful* iteration.

In other words, Δ_k provides a *certificate* of stationarity:

- Either $\| \nabla f(x_k) \|$ is on the order of Δ_k,
- or the function is so ill-behaved that accurate identification of a stationary point is difficult without the use of curvature (second-order) information.

Finally:

All these ideas can be extended to handle constraints.

We replace $\| \nabla f(x_k) \|$ with an appropriate measure of constrained stationarity.

We now require $\mathcal{D}_k$ to contain generators for cones defined by the nearby constraints (or linearizations of the nearby constraints).

Once again, we show that there exists a relationship between Δ_k and the measure of stationarity at unsuccessful iterations.

From this, we then derive global convergence results, as well as obtaining a certificate of constrained stationarity.

For more details:

Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods, Kolda/Lewis/Torczon, SIAM Review (45) 2003, pp. 385–482.

Stationarity results for generating set search for linearly constrained optimization, Kolda/Lewis/Torczon, revised July 2004.

`www.cs.wm.edu/~va/research`