Stationarity results for generating set search for linearly constrained optimization

Virginia Torczon, College of William & Mary

Collaborators:

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

The problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax \leq b, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$, A is an $m \times n$ matrix, and $b \in \mathbb{R}^m$.

We use Ω to denote the feasible region: $\Omega = \{ x \in \mathbb{R}^n : Ax \leq b \}$.

We assume that f is continuously differentiable on Ω but that gradient information is not computationally available.

We do not assume that the constraints are nondegenerate.

The stationarity measure:[1]

$$\chi(x) \equiv \max_{x+w\in\Omega,\ \| w \|\leq 1} -\nabla f(x)^T w,$$

where $\chi(x)$ is a continuous function with the property

$$\chi(x) = 0 \text{ for } x \in \Omega$$

if and only if

x is a KKT point of the linearly constrained problem.

---

[1] A. R. Conn, N. Gould, A. Sartenaer, and P. L. Toint, *Convergence propreties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints*, SIOPT, 1996.

First-order stationarity

Main result: Once Δ_k is small enough the subsequence x_k for $k \in \mathcal{U} \subseteq \{0, 1, 2, \ldots\}$ produced by a Generating Set Search (GSS) method for the linearly constrained optimization problems satisfies

$$\chi(x_k) = O(\Delta_k).$$

Observe: a careful specification of a GSS method for linearly constrained optimization ensures, at a minimum, that $\liminf_{k \to \infty} = 0$, so that first-order stationarity for this subsequence of iterates is immediate.

But that is *not* the primary motivation for this investigation since global convergence to KKT points for the linearly constrained problems have already been established [see May, 1974; Yu/Li, 1981; Lewis/Torczon, 2000; Lucidi/Sciandrone/Tseng, 2002;. . ..]

What are GSS methods for linearly constrained optimization?

Look at one simple example applied to the problem:

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

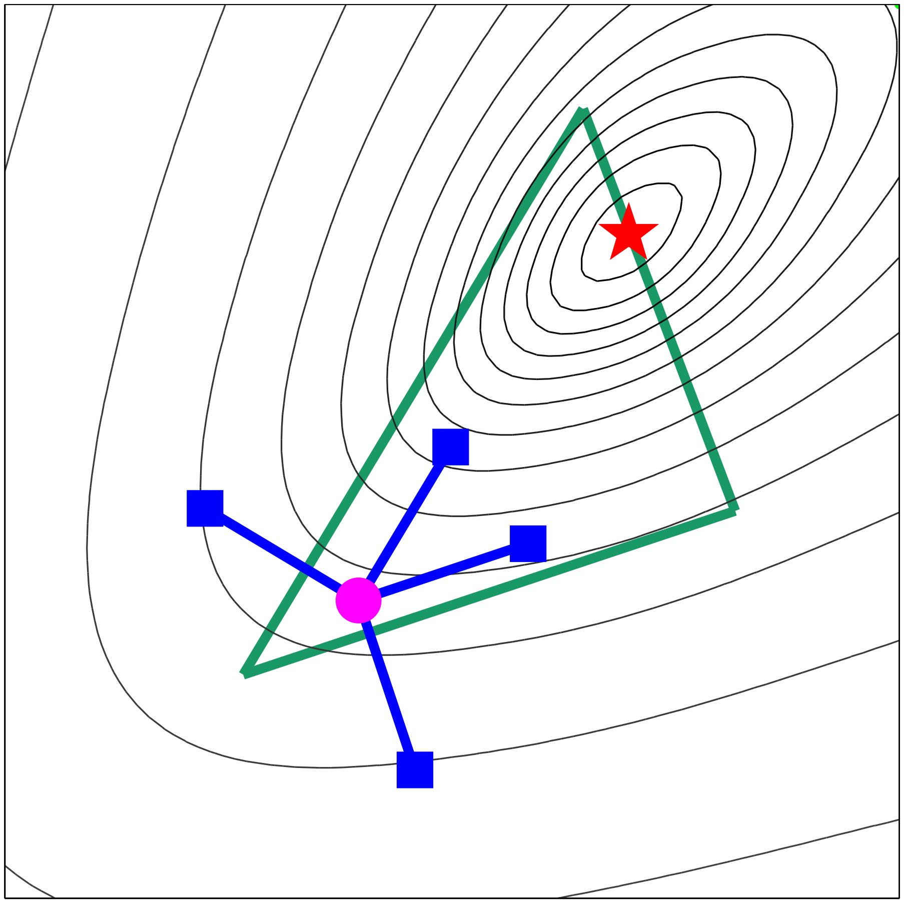
where

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

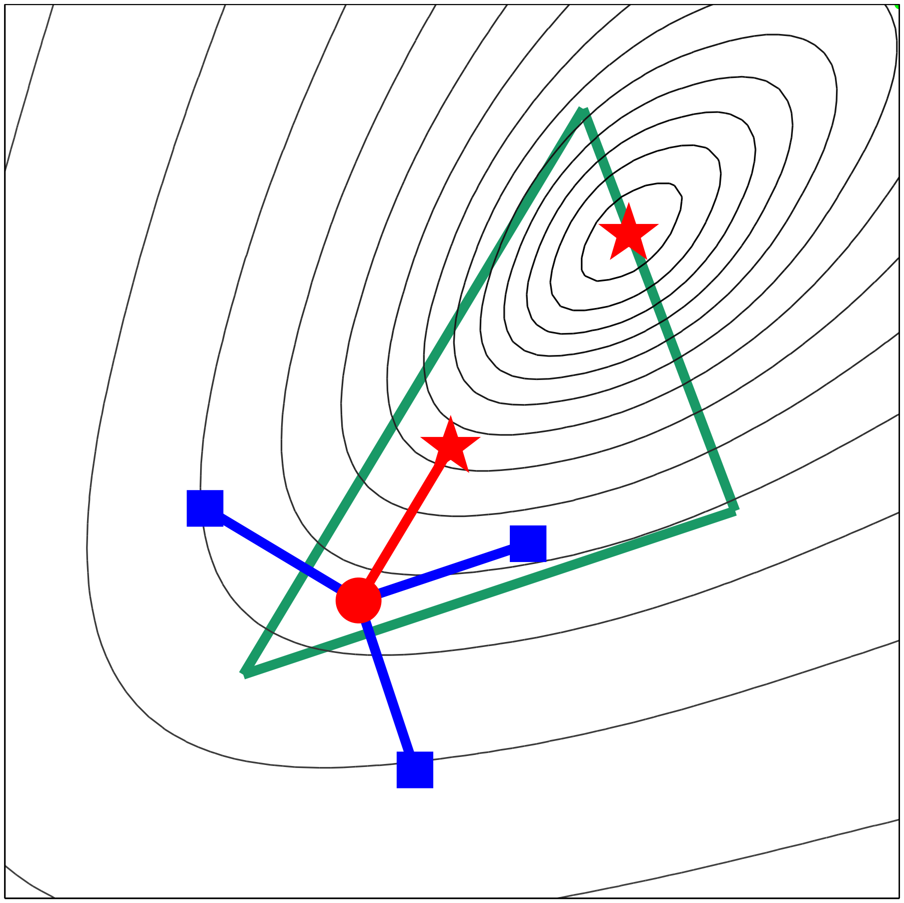
—the modified Broyden tridiagonal function—augmented with three linear constraints.

Use a *feasible iterates* approach.

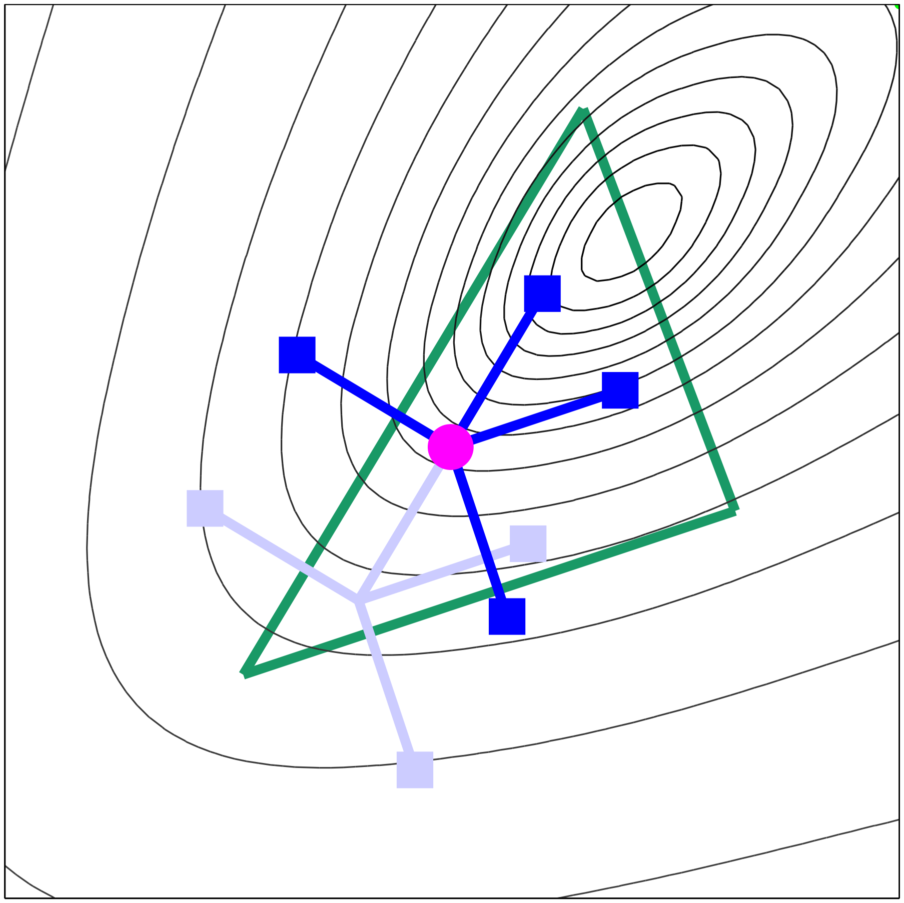
The initial set of search directions:



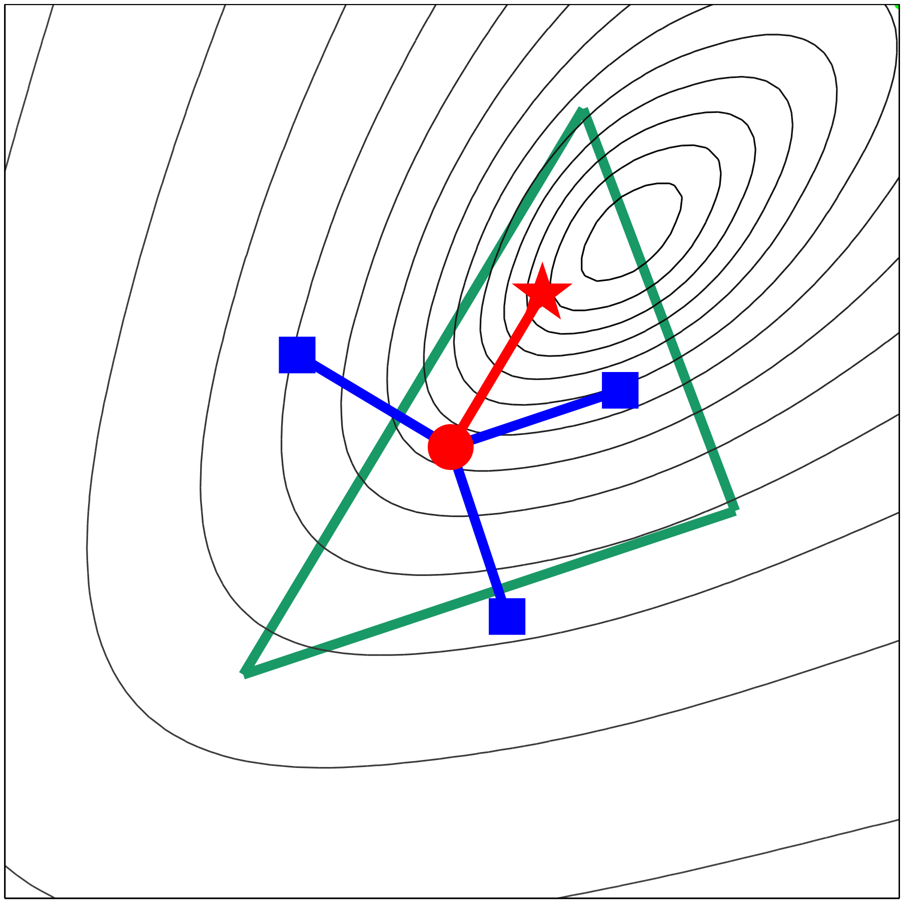
Identify feasible improvement:



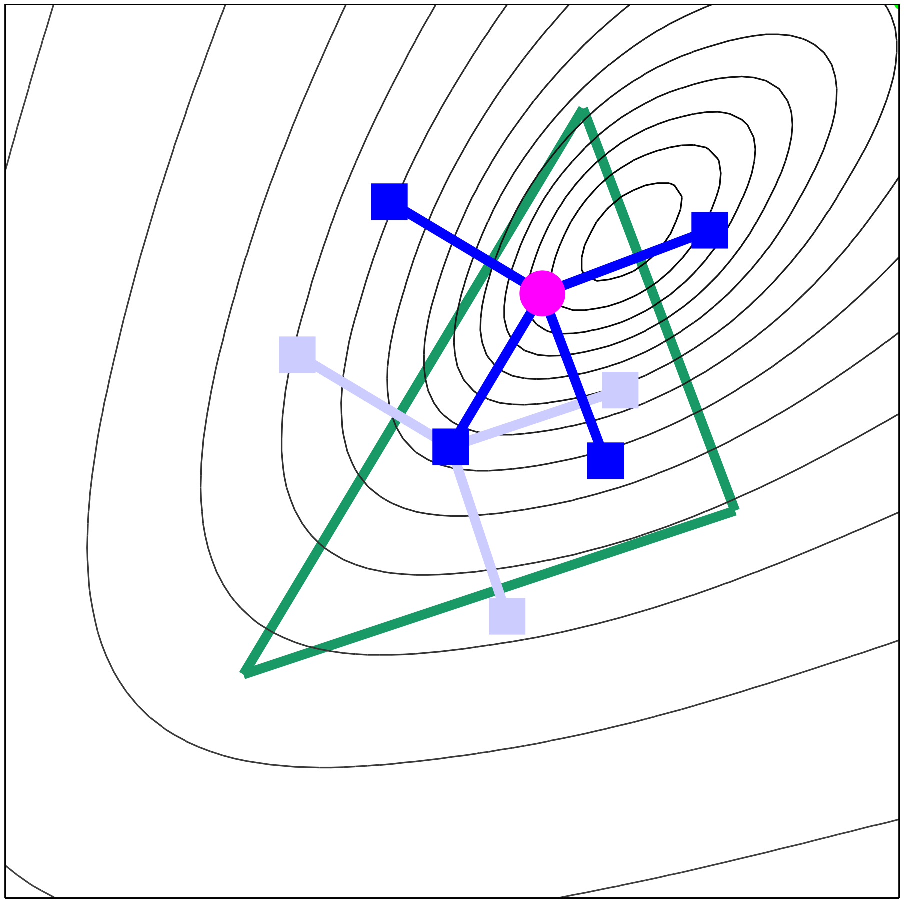
Move Northeast and keep the set of search directions



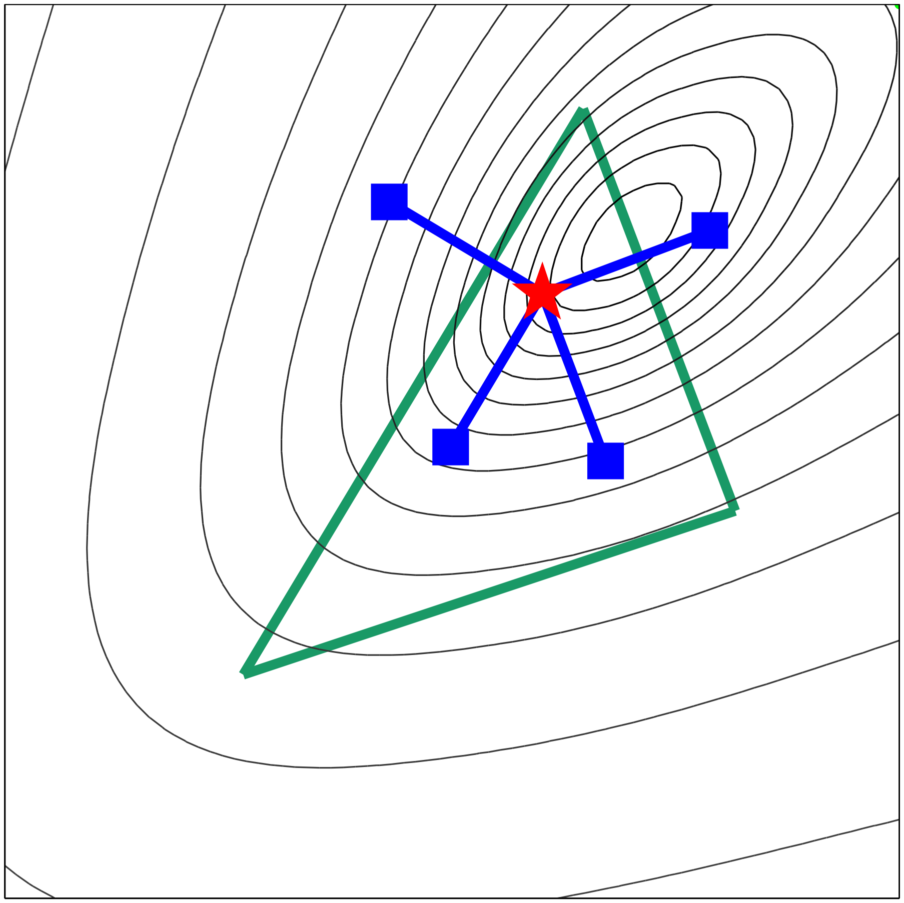
Identify feasible improvement:



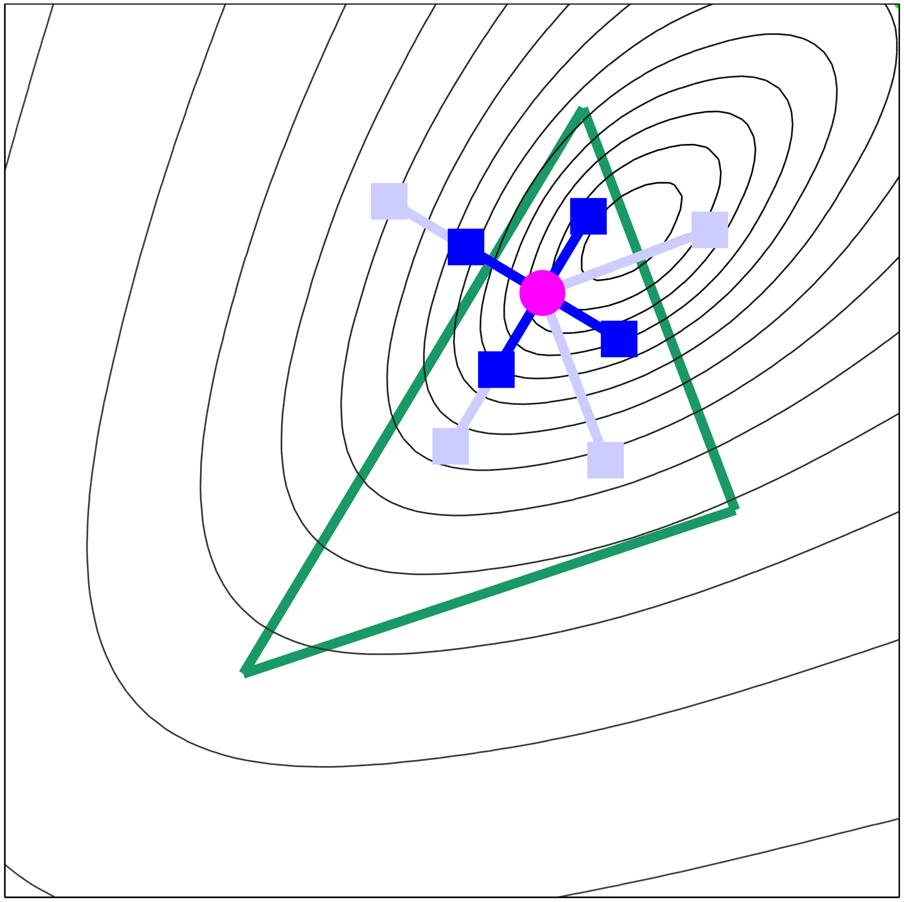
Move Northeast and *change* the set of search directions



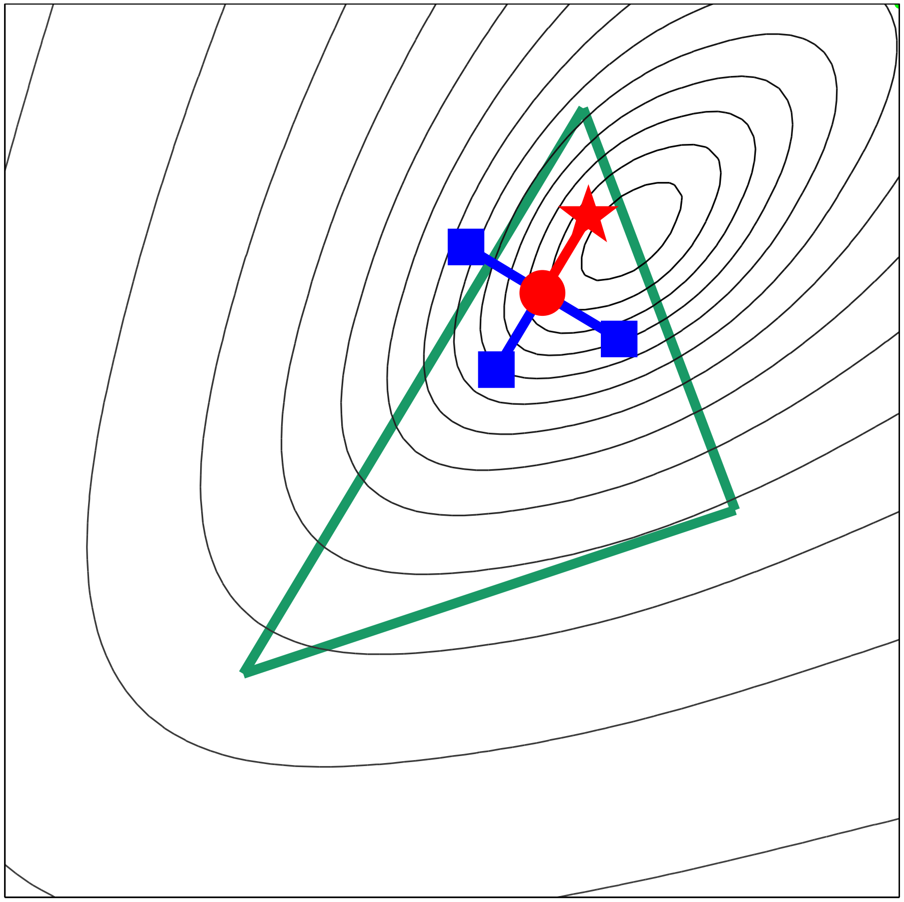
No feasible improvement:



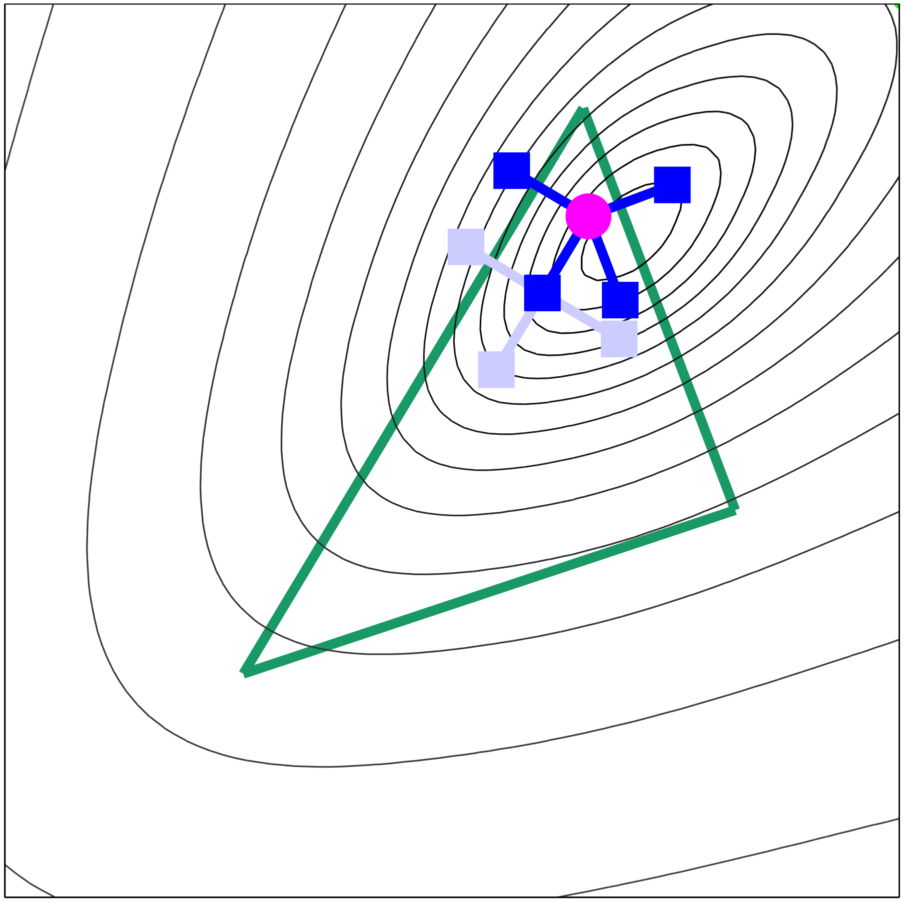
Contract and change the set of search directions



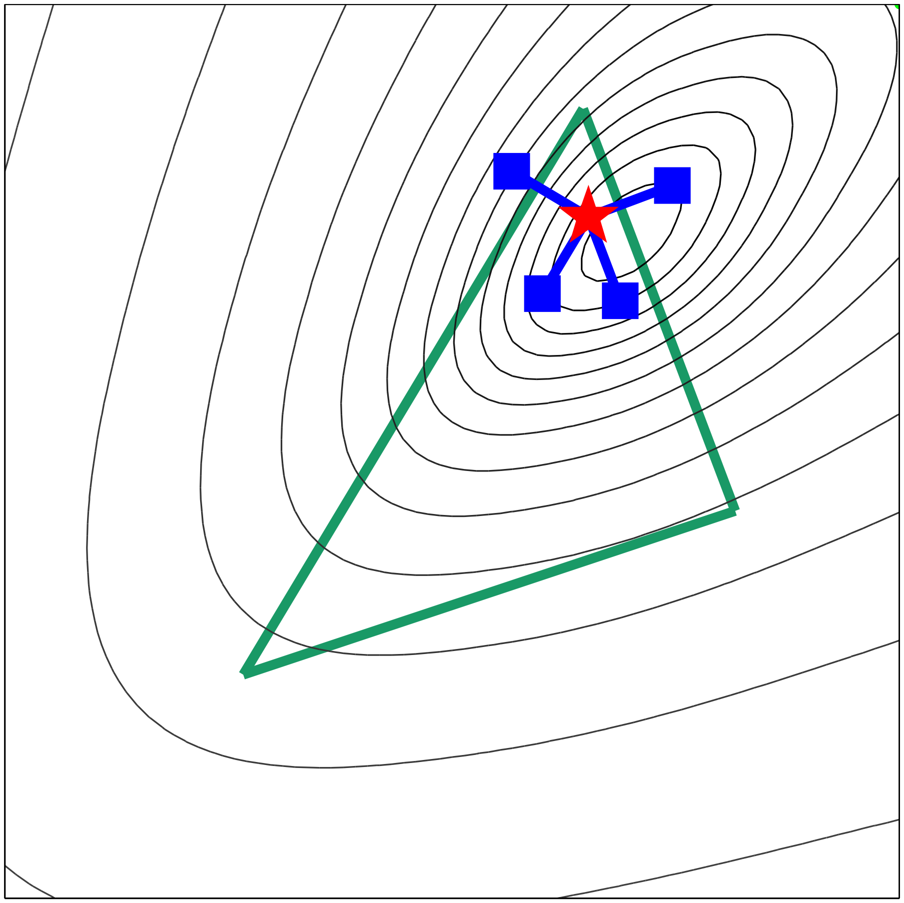
Identify feasible improvement:



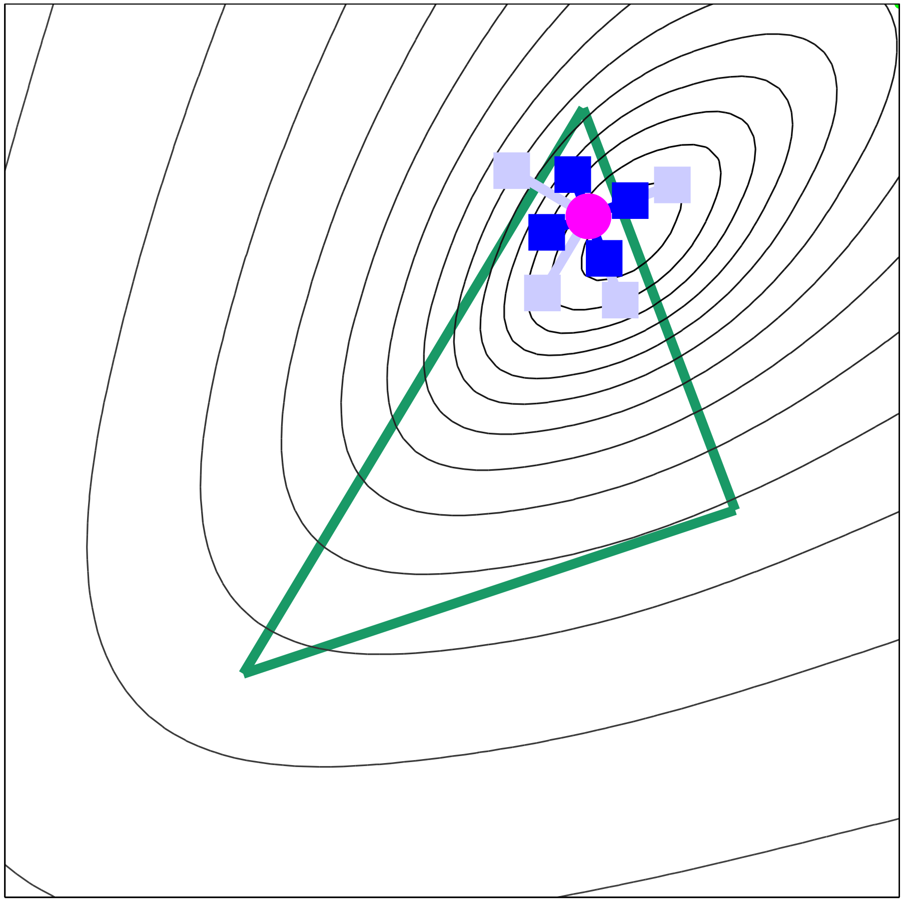
Move Northeast and *change* the set of search directions



No feasible improvement:



Contract and change the set of search directions



Critical:

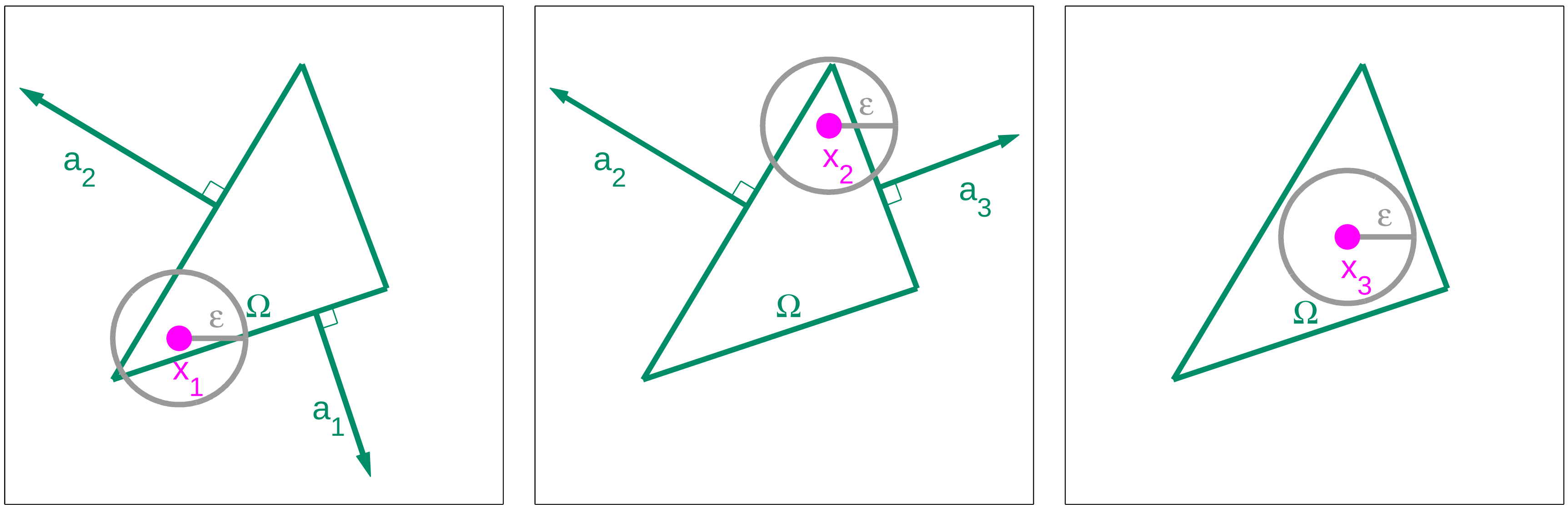
When *close* to the boundary of the feasible region, the set of directions must conform to the geometry of the nearby constraints.

Essential features:

- identifying the nearby constraints
- obtaining a set of search directions
- finding a step of an appropriate length
- accepting a step

Identifying the nearby constraints

Find the outward-pointing normals within distance ε of the current iterate.

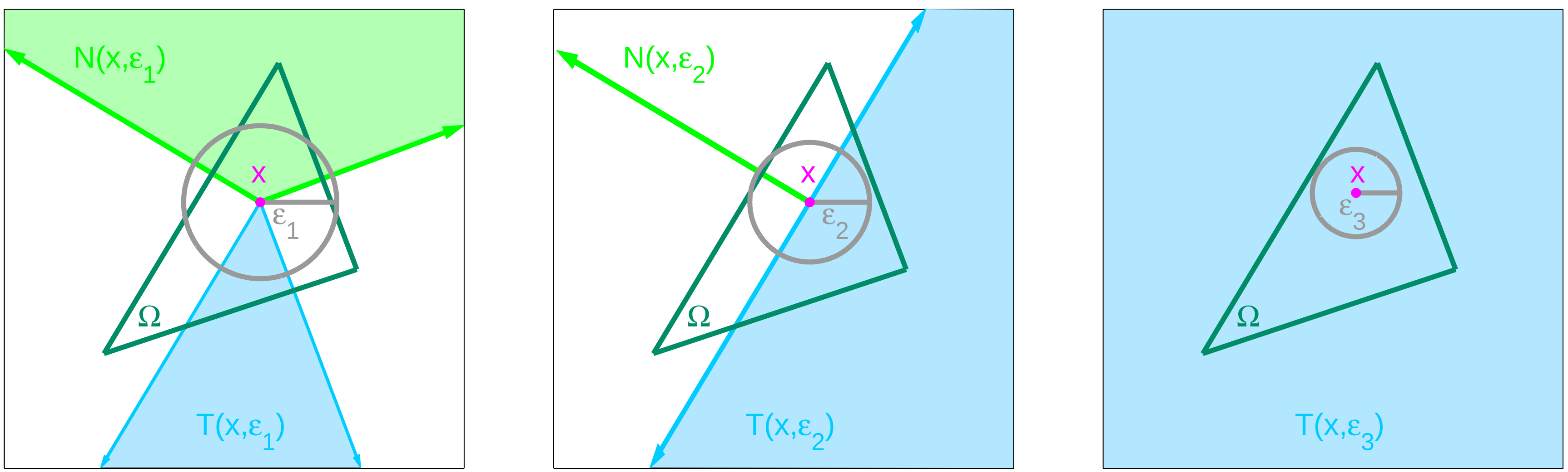


The conditions on ε depend on the convergence analysis in effect.

Obtaining a set of search directions: Part I

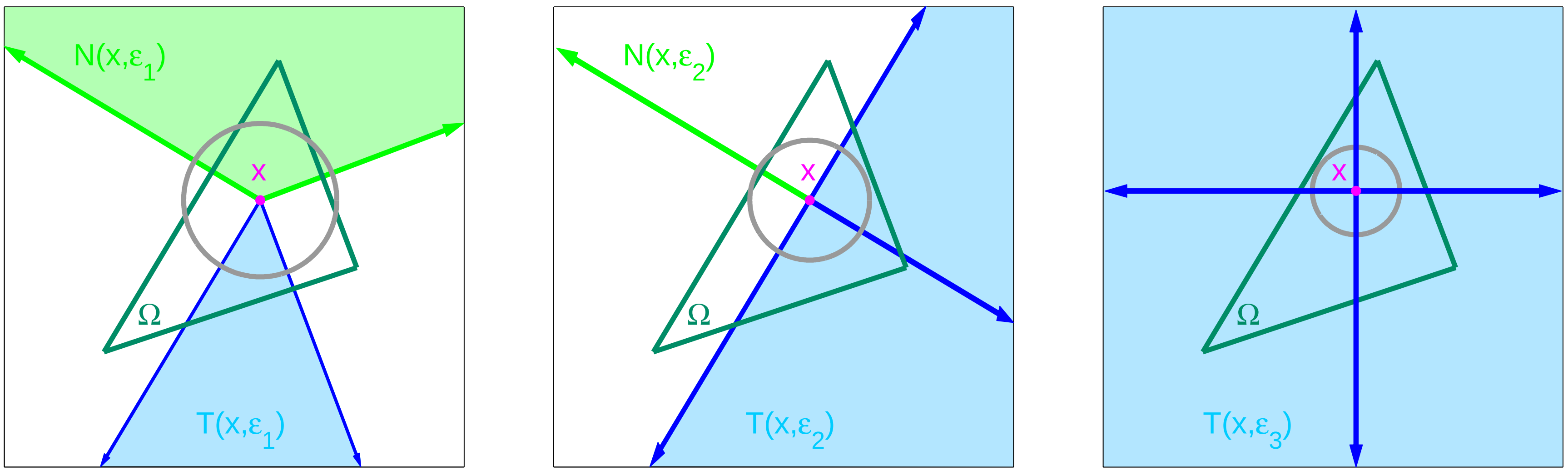
Translate the outward-pointing normals within distance ε of the current iterate x to obtain

- the ε-*normal cone* $N(x, \varepsilon)$ and
- its polar, the ε-*tangent cone* $T(x, \varepsilon)$.



Obtaining a set of search directions: Part II

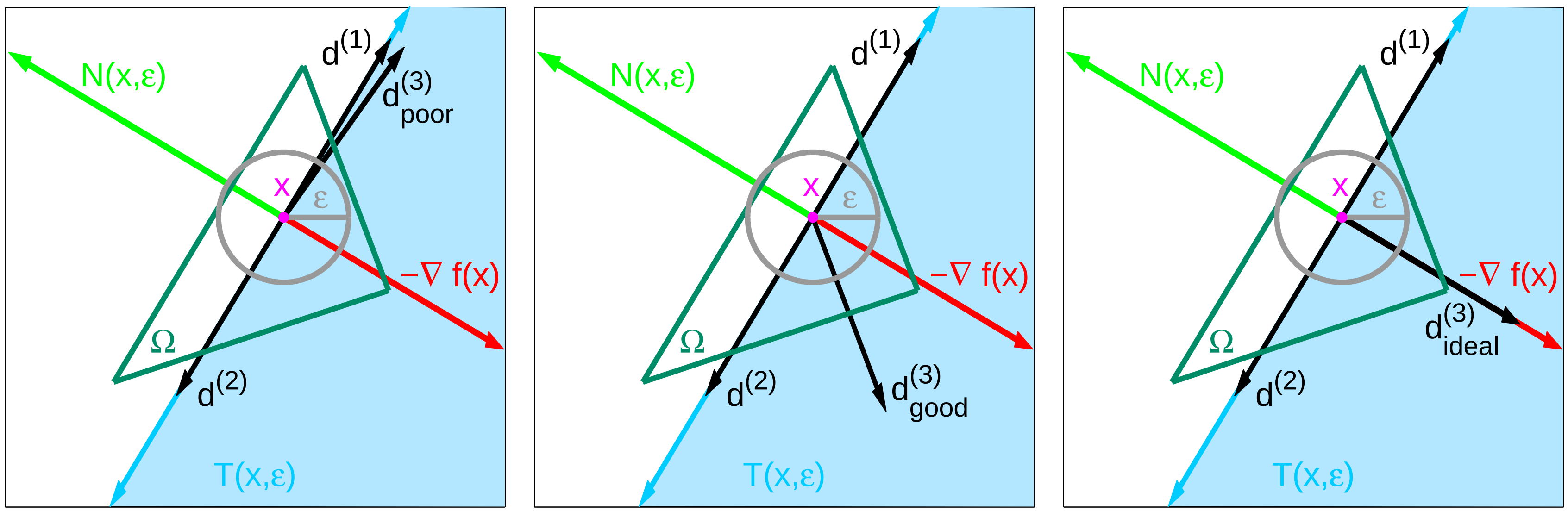
The set of search directions *must* contain generators for the ε-tangent cone $T(x, \varepsilon)$.



Conditioning is important!

Conditioning: concern

When $T(x,\varepsilon)$ is a lineality space or half space, there is freedom in choosing the generators:



Conditioning: requirement that must be enforced

There exists a constant $\kappa_{\min} > 0$, independent of k, such that for all k there exists a set of generators $\mathcal{G}$ for $T(x_k, \varepsilon_k)$ and, furthermore,

$$\kappa(\mathcal{G}) \equiv \min_{\substack{v \in \mathbb{R}^n \\ v_K \neq 0}} \max_{d \in \mathcal{G}} \frac{v^T d}{\| v_K \| \| d \|} \geq \kappa_{\min}.$$

Conditioning: gratis

There exists a constant $\nu_{\min} > 0$, independent of k, such that for all k there exists a set of generators $\mathcal{A}$ for $N(x_k, \varepsilon_k)$ such that

$$\kappa(\mathcal{A}) \geq \nu_{\min}.$$

The "for free" is because each $N(x, \varepsilon)$ is generated by at most m rows of the constraint matrix A.

Obtaining a set of search directions: Part III

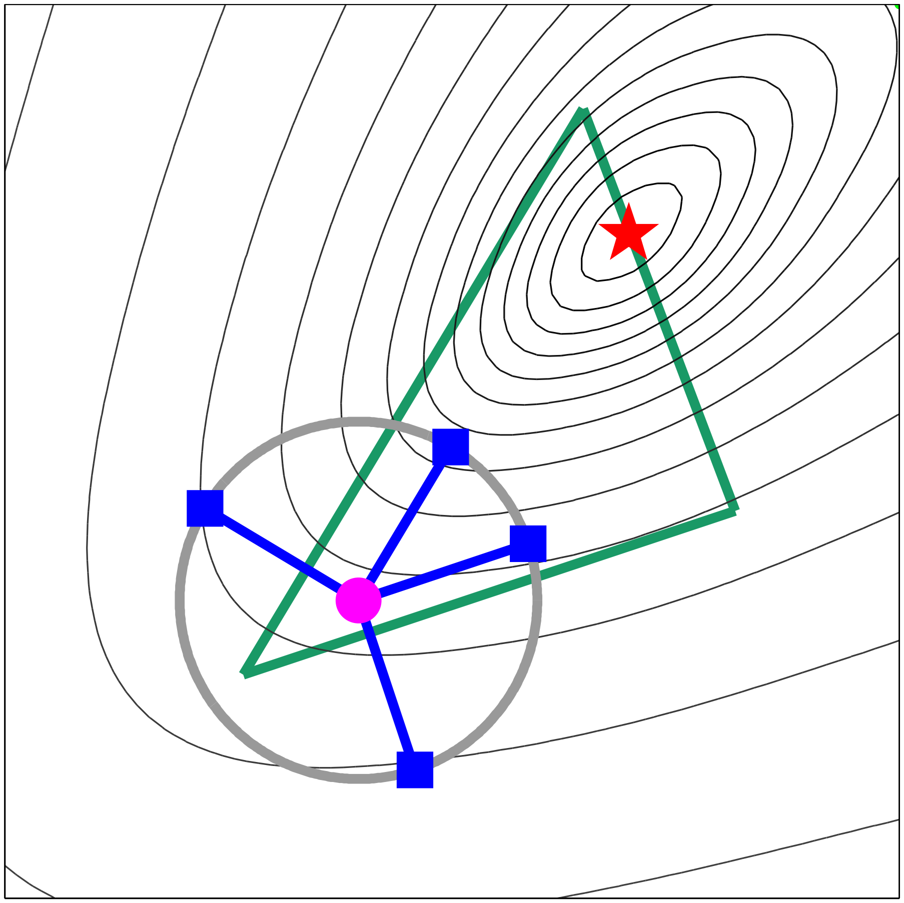
What—*precisely*—the set of search directions contains depends on the convergence analysis in effect.

Choices:

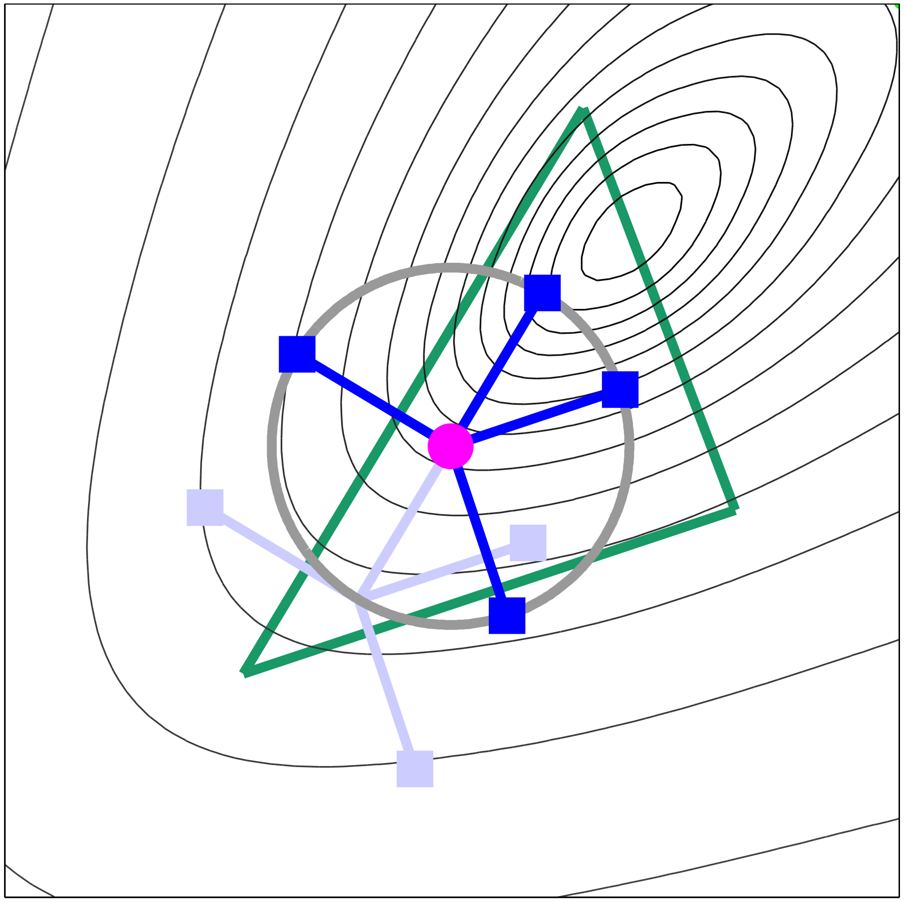
- At a minimum, generators for the ε-tangent cones $T(x_k, \varepsilon)$ for all $\varepsilon \in [0, \varepsilon_*]$. [Lewis/Torczon, 2000]
- *Only* generators for the ε-tangent cone $T(x_k, \varepsilon_k)$. [Lucidi/Sciandrone/Tseng, 2002]
- At a minimum, generators for the ε-tangent cone $T(x_k, \varepsilon_k)$, where $\varepsilon_k = \min\{\varepsilon_{\max}, \beta_{\max}\Delta_k\}$. [Kolda/Torczon/Lewis, 2004]

The sets used for the example contained generators for *both* $T(x_k, \varepsilon_k)$ and $N(x_k, \varepsilon_k)$.

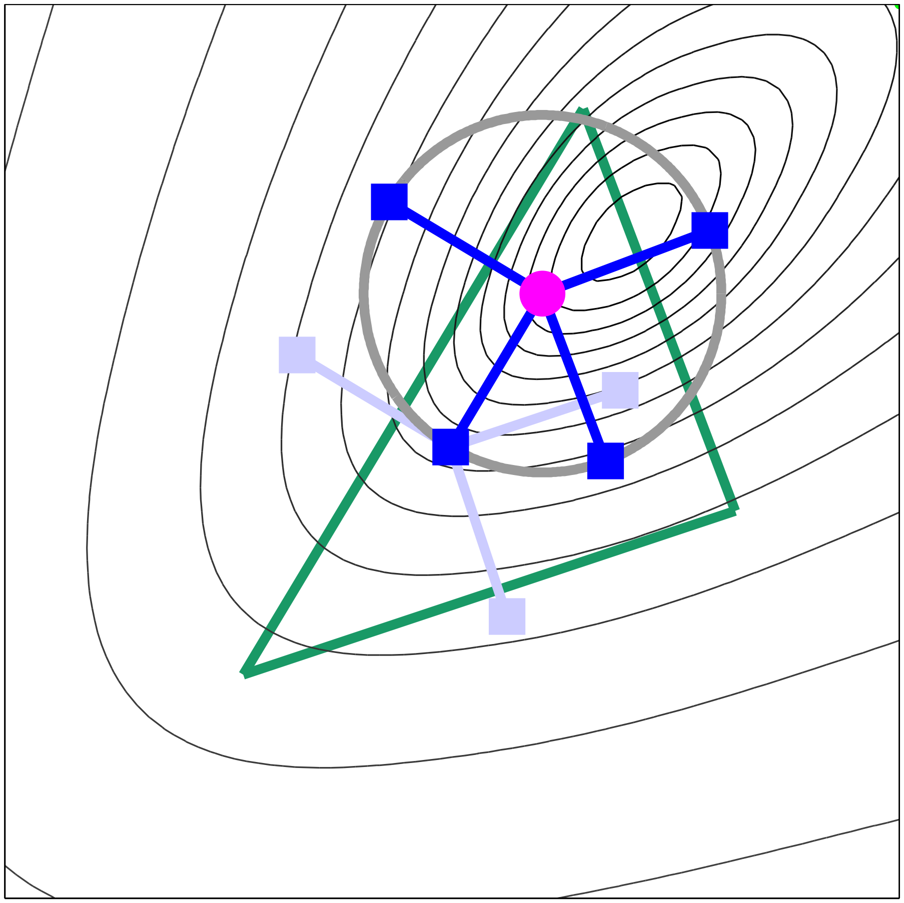
The initial set of search directions:



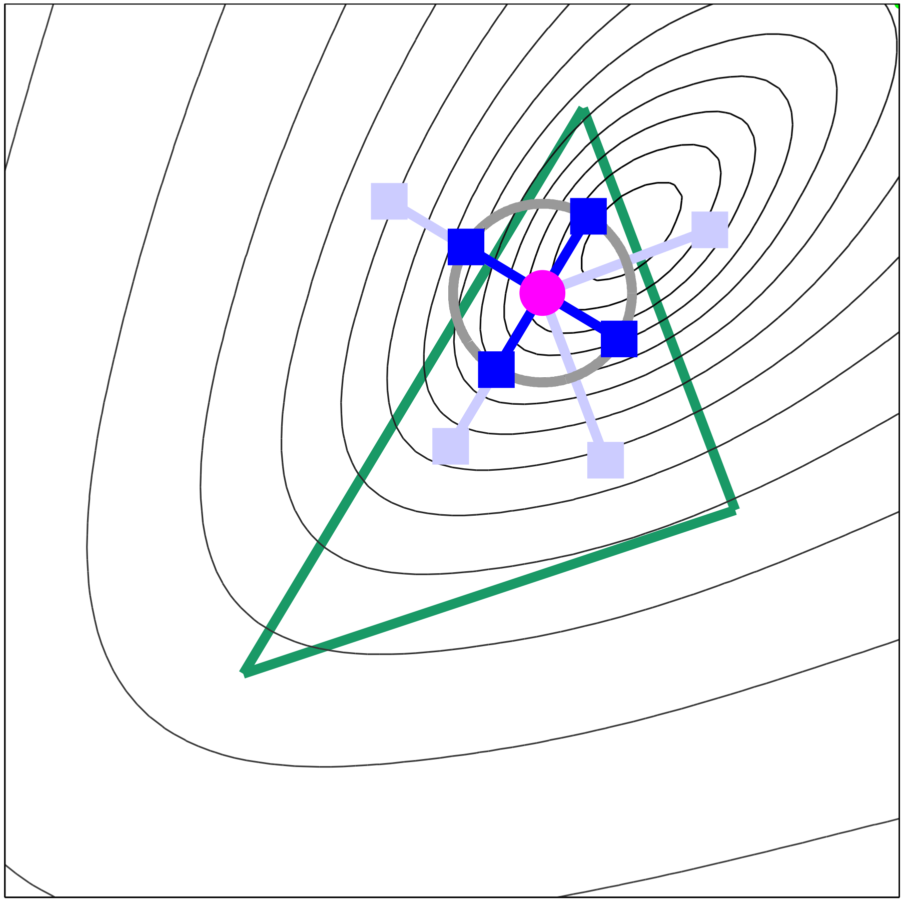
Move Northeast and keep the set of search directions



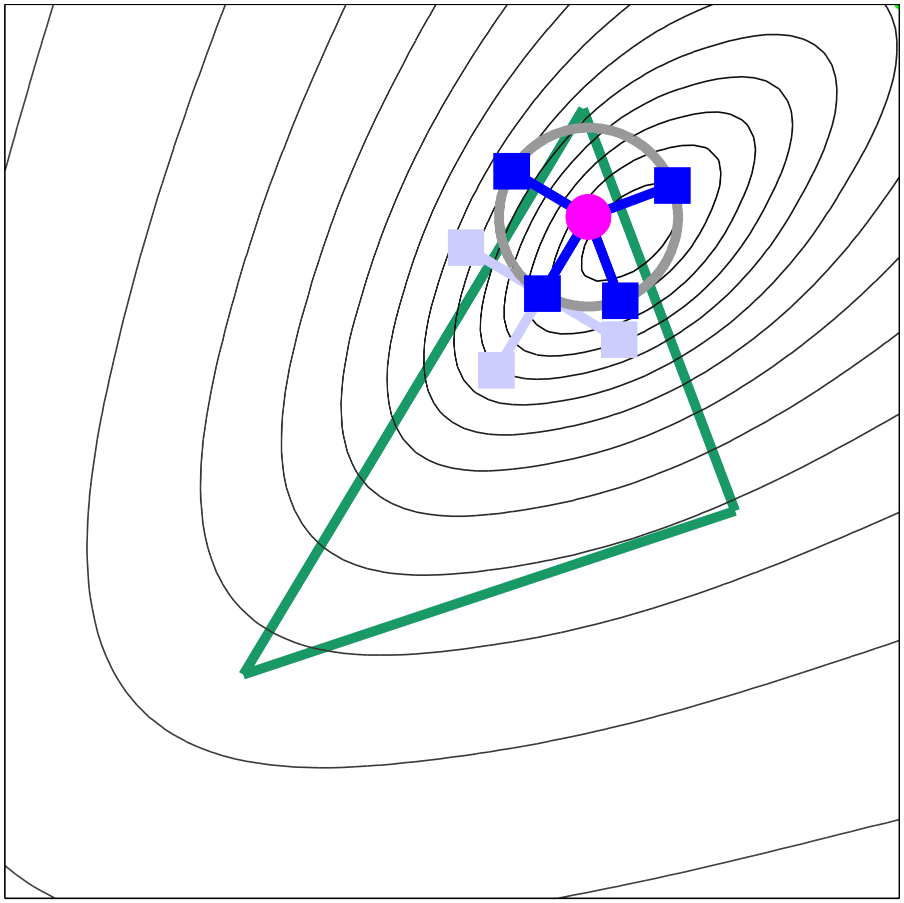
Move Northeast and *change* the set of search directions



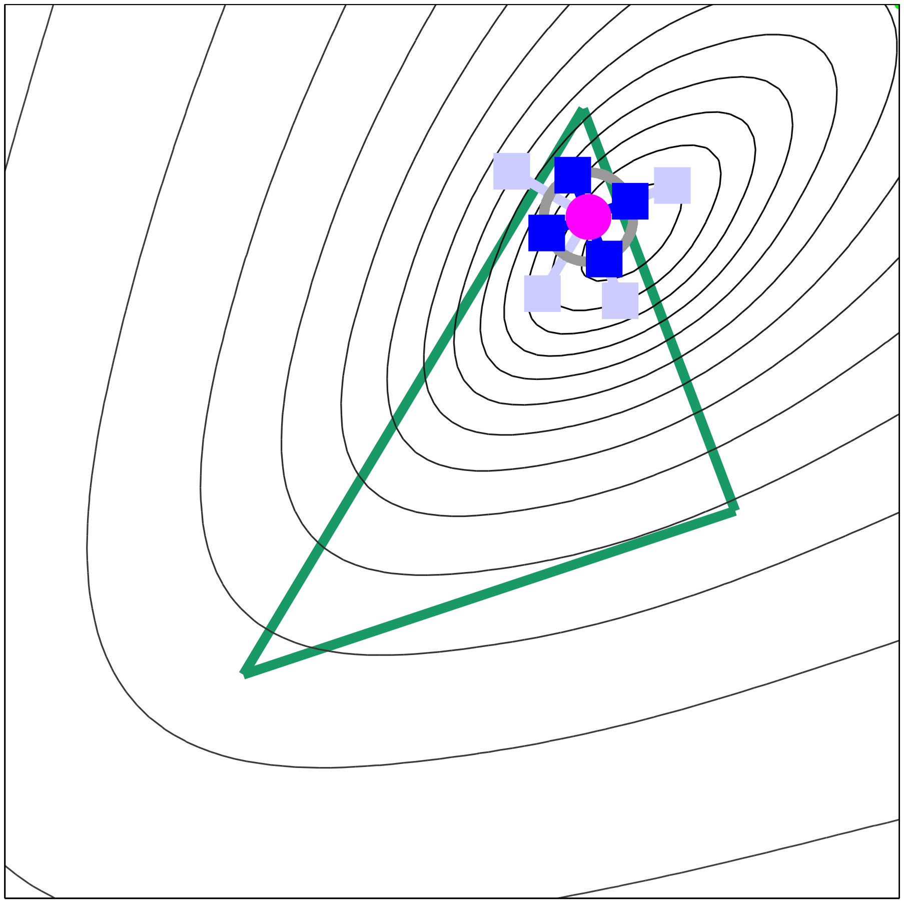
Contract and change the set of search directions



Move Northeast and *change* the set of search directions



Contract and change the set of search directions



Finding steps of an appropriate length

First: bound the lengths of the direction vectors.

Specifically, there must exist $\beta_{\min}$ and $\beta_{\max}$, independent of k, such that for all k the following holds:

$$\beta_{\min} \leq \| \, d \, \| \leq \beta_{\max} \quad \text{for all } d \in \mathcal{G}_k.$$

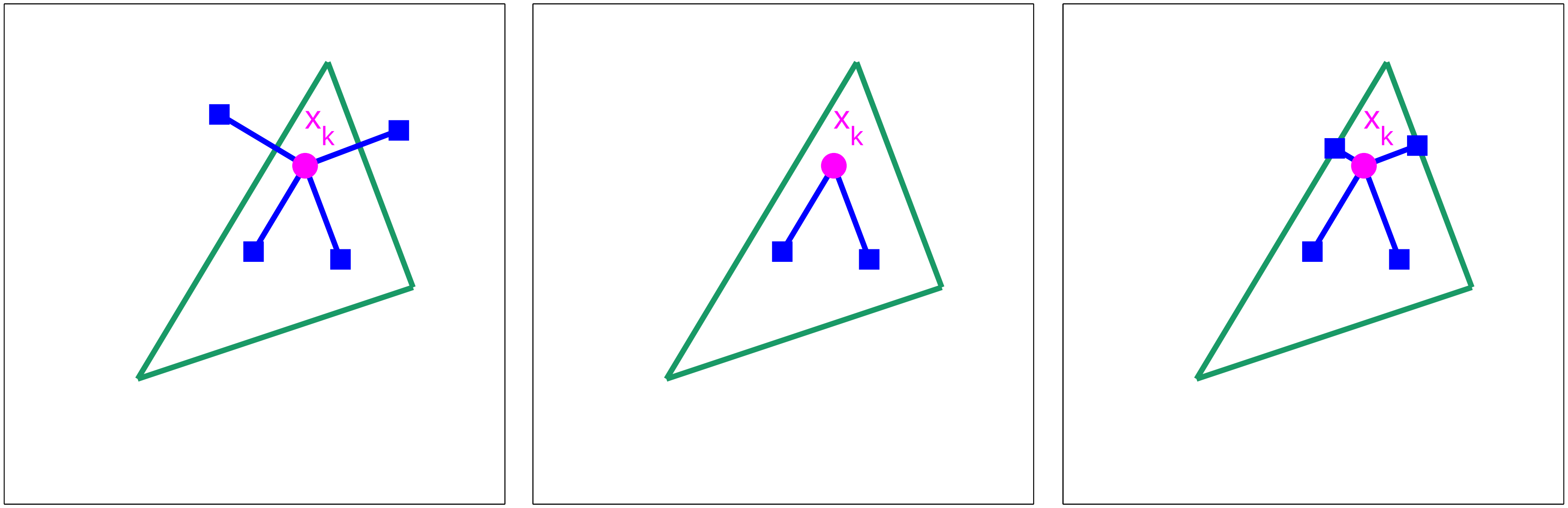
Finding steps of an appropriate length

Second: once the lengths of the direction vectors are bounded for all k, tie the lengths of the steps tried to a step-length control parameter Δ_k.

Requirement: If $x_k + \Delta_k d_k^{(i)} \in \Omega$, then the actual step taken along $d^{(i)}$ must be of length Δ_k, just as for unconstrained generating set search methods.

Question: What to do when $x_k + \Delta_k d_k^{(i)} \notin \Omega$?

Finding *feasible* steps of an appropriate length



Once again, the analysis allows multiple options.

Accepting a step

If

$$x_k + \tilde{\Delta}_k d_k \in \Omega$$

and

$$f(x_k + \tilde{\Delta}_k) < f(x_k) - \rho(\Delta_k)$$

then $k \in \mathcal{S}$,

$$x_{k+1} = x_k + \tilde{\Delta}_k$$

and

$$\Delta_{k+1} = \phi_k \Delta_k \text{ for a choice of } \phi_k \geq 1.$$

Otherwise, $k \in \mathcal{U}$,

$$x_{k+1} = x_k,$$

and

$$\Delta_{k+1} = \theta_k \Delta_k, \text{ for some choice } \theta_k \in (0, 1).$$

The forcing function:

1. The function $\rho(\cdot)$ is a nonnegative continuous function on $[0, +\infty)$.

2. The function $\rho(\cdot)$ is $o(t)$ as $t \downarrow 0$; i.e., $\lim_{t \downarrow 0} \rho(t)/t = 0$.

3. The function $\rho(\cdot)$ is nondecreasing; i.e., $\rho(t_1) \leq \rho(t_2)$ if $t_1 \leq t_2$.

The stationarity result:

If the set $\mathcal{F} = \{x \in \Omega \,|\, f(x) \leq f(x_0)\}$ is bounded and the gradient of f is Lipschitz continuous with constant M on $\mathcal{F}$, then there exists $\gamma > 0$ such that for all $x \in \mathcal{F}$, $\| \nabla f(x) \| < \gamma$.

Further, for all $k \in \mathcal{U}$, if $\varepsilon_k = \beta_{\max} \Delta_k$, then

$$\chi(x_k) \leq \left(\frac{M}{\kappa_{\min}} + \frac{\gamma}{\nu_{\min}} \right) \Delta_k \beta_{\max} + \frac{1}{\kappa_{\min} \beta_{\min}} \frac{\rho(\Delta_k)}{\Delta_k}.$$

Conclusions:

- Δ_k can be used to assess progress toward a KKT point.

- The bound on $\chi(x_k)$ illuminates what algorithmic parameters can—and should—be monitored to assure the effectiveness of an implementation.

- Our analysis yields an estimate that makes it possible to use linearly constrained GSS methods with the augmeted Lagrangian approach of Conn/Gould/Sartenaer/Toint to handle problems with general constraints.