Most of iterative techniques utilize a projection process. The main idea is to extract an approximate solution from a subspace. If \( K \) is the search subspace of dimension \( M \), then, in general, \( M \) constraints must be imposed to be able to extract such an approximation. These constraints (set of \( M \) conditions) are typically expressed by orthogonally constraints on the residual vector. We call \( L \) the subspace of constraints.

This framework is common to many different mathematical methods and is known as the "Petrov - Galerkin" condition.

\[ \text{If } L = K \Rightarrow \text{orthogonal projection} \]
\[ \text{If } L \neq K \Rightarrow \text{oblique projection} \]
Problem

Let $A \in \mathbb{R}^{n \times n}$ solve $Ax = f$

Find approximate solution $x \in K$ such that

$\hat{r} = f - Ax \perp L$

$[v_1, v_2, \ldots, v_m] = V_{n \times m} \rightarrow$ basis for $K$

$[w_1, w_2, \ldots, w_n] = W_{n \times n} \rightarrow$ basis for $L$

Let $x_0$ be an initial iterate and

$\hat{x} = x_0 + \lambda y$

$\hat{r} = f - A\hat{x} = f - A(x_0 + \lambda y)$

$\hat{r} = r_0 - AVy$

with condition $W^T_t = 0$

$\Rightarrow W^T_t (r_0 - AVy) = 0 \Rightarrow (W^T AV)y = W^T r_0$

Solve for $y$ (small linear system $\Rightarrow$ projected problem)

$\Rightarrow [\hat{x} = x_0 + V(W^T AV)^{-1} W^T r_0]$
Prototype Projection Method

Do:
1. Select a pair of subspaces \( K \) and \( L \)
2. Choose bases \( V \) and \( W \)
3. Compute
   \[ f - A x = \begin{pmatrix} w^T A v \\ w^T \end{pmatrix} \]
4. Solve
   \[ y = (w^T A v)^{-1} w^T \]
5. Update
   \[ x \leftarrow x + V y \]

P. Approx. solution is defined only when \((W^T A V)\)
   is non-singular.

Theorem

Let \( A, K \) and \( L \) satisfy either one of the following conditions:

\[ \begin{cases} 
(i) & A \text{ is singular and } L = K \\
(ii) & A \text{ is non-singular and } L = AK \\
\end{cases} \]

Then \( B = W^T A V \) is non-singular for any bases \( V \) and \( W \) of \( K \) and \( L \).
A particular class of projection methods is the one-dimensional projection process (single vector):

\[ \begin{align*}
K &= \text{span}(\{v\}) \\
L &= \text{span}(\{w\})
\end{align*} \]

Scheme: Do until convergence

\[ x = x - Ax \]

\[ d = \frac{w^T r}{w^T Av} \]

\[ x \leftarrow x + \alpha d \]

End do

3 regular choices

- A spd \( \rightarrow \) Steepest descent \( \Rightarrow V = r, W = r \)
- A^T A \( \text{ spd } \) \( \rightarrow \) Minimal Residual direction \( \Rightarrow V = r, W = Av \)
- Residual norm Steepest descent \( \rightarrow V = A^T r, W = Av \)

L is equivalent to Steepest descent applied to normal equation

\[ A^T A x = A^T r \]
Steeped descent

\[ A = \text{spd}, \quad v = 1, \quad w = 1 \]

\[ \begin{align*}
\text{Iteration:} & \quad x = x - \frac{f - Ax}{r^T A r} \\
& \quad x \leftarrow x + \alpha r
\end{align*} \]

(Convergence guaranteed if \( A \) is spd)

\[ \text{Requires only matrix-vector product + inner product} \]

Each step minimizes the functional (ex act solution)

\[ \begin{align*}
\psi(x) &= \| x - x^* \|^2_A = (x - x^*)^T A (x - x^*)
\end{align*} \]

One can think of the method of steepest descent from a different perspective — as a method that belongs to the much larger class of "Gradient Methods"

\[ \begin{align*}
\rightarrow \text{Find a local minimum of the function } \psi \text{ using gradient descent: } \\
\psi(x) &= \frac{1}{2} x^T A x - f x
\end{align*} \]

By: "Converting a linear problem into a quadratic one"

\[ \rightarrow \text{Find minimum of } \psi(x) \text{ instead of solving } A x = f \]
S. Puig \(Ax = f\)  

Minimize \(Q(x) = \frac{1}{2}x^T A x - b^T x + c\)  

Indeed \(\text{gradient of } Q(x)\) is given by  
\[\nabla Q(x) = \frac{1}{2} (A + A^T) x - b = A x - b\]  

\[\frac{1}{2} (A + A^T) = A \text{ because } A \text{ is symmetric}\]  

Since \(A\) is spd, quadratic form \(Q(x)\) is strictly convex  
\[\Rightarrow \text{unique minimizer } x \text{ given by } \nabla Q(x) = 0\]  

The surfaces \(Q(x) = \text{constant}\) (level contour set) form a family of ellipsoids with a common center \(x = A^{-1} f\)  

Example  
2\(\text{copositive } x = \frac{1}{2} x\)
Steepest descent

Starting with step $k$, $x_k$, go in direction of the steepest descent of $\nabla f(x)$ (direction where $f(x)$ is decreasing most rapidly) equaled by $-\nabla f(x)$.

$$x_{k+1} = x_k + \alpha_k \nabla f(x_k)$$

**Stepsize**

$$\alpha_k = \frac{1}{A}$$

We expect that taking a step in that direction will bring us closer to the minimum (i.e., line search).

**Question**: What step size to use at each iteration? i.e., value of $\alpha_k$?

Logical approach $\Rightarrow x_{k+1} = x_k + \alpha_k \nabla f(x)$ minimizes $f(x_{k+1})$

$$\nabla f(x_{k+1}) = 0$$

Then

$$\alpha_k = \frac{1/T}{\| \nabla f(x) \|^2}$$

After calculation.
By: Steepest descent may "Zigzag" too much.
   ⇒ Very tiny step towards the end (due to convergence)

Example

The conjugate gradient (CG) technique removes this issue.

Conjugate Gradient

One of the best known iterative techniques for
solving SPD system ⇒ highly efficient.

In contrast to Steepest descent, CG uses a
search direction that is linearly independent to all
previous search directions.

Search directions are \( A \) orthogonal. \( d_t^T A d_t = 0 \) (≠ 0)
⇒ Conjugate
x It is mathematically equivalent to full orthogonalization method (FOM).

\[x\] exact solution reached in at most

\[N\] steps (since \(N\) linearly ind. search directions)
will span the whole space \(\Rightarrow\) direct search true

\(\Rightarrow\) it is equivalent to an orthogonal projection technique
onto the Krylov subspace (see next section)

\(\Rightarrow\) convergence can be accelerated by preconditioning.

\(\) Alg.

\(x^0 \) initial guess.

\[x^k = x^k - Ax^k\]

\[d^k = r^k\] search direction initial

\(\Rightarrow\) Do until convergence.

\[d^k = \frac{r^k \times r^k}{d^k A d^k}\]

\[x^{k+1} = x^k + d^k d^k\] new iterate.

\[r^{k+1} = r^k - d^k A d^k\] new residual.

\[p^{k+1} = \frac{r^{k+1} r^{k+1}}{r^k r^k}\] compute new direction.

\[d^{k+1} = r^{k+1} + p^{k+1} d^k\]

\(\Rightarrow\) end do.