Conjugate Directions

- Powell's method is based on a model quadratic objective function and conjugate directions in $\mathbb{R}^n$ with respect to the Hessian of the quadratic objective function.

- what does it mean for two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ to be conjugate ?

Definition: given that $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, then $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *mutually orthogonal* if $(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T\boldsymbol{v} = 0$ (where $(\boldsymbol{u}, \boldsymbol{v})$ is our notation for the *scalar product*).❐

Definition: given that $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, then $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *mutually conjugate* with respect to a symmetric positive definite matrix A if $\boldsymbol{u}$ and $A\boldsymbol{v}$ are mutually orthogonal, *i.e.* $\boldsymbol{u}^T A\boldsymbol{v} = (\boldsymbol{u}, A\boldsymbol{v}) = 0$.❐

- Note that if two vectors are mutually conjugate with respect to the identity matrix, that is $A = I$, then they are mutually orthogonal.

Eigenvectors

- $\boldsymbol{x}_i$ is an eigenvector of the matrix A, with corresponding eigenvalue λ_i if it satisfies the equation

$$A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i \quad i = 1, \ldots, n$$

  and λ_i is a solution to the characteristic equation $|A - \lambda_i I| = 0$.

- If $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, then there will exist n eigenvectors, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ which are mutually orthogonal (*i.e.* $(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$ for $i \neq j$).

- Now since: $(\boldsymbol{x}_i, A\boldsymbol{x}_j) = (\boldsymbol{x}_i, \lambda\boldsymbol{x}_j) = \lambda(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$ for $i \neq j$, this implies that the eigenvectors, $\boldsymbol{x}_i$, are mutually conjugate with respect to the matrix A.

We Can Expand Any Vector In Terms Of A Set Of Conjugate Vectors

Theorem: A set of n mutually conjugate vectors in $\mathbb{R}^n$ span the $\mathbb{R}^n$ space and therefore constitute a basis for $\mathbb{R}^n$.❐

Proof:

let $\boldsymbol{u}_i$, $i = 1, \ldots, n$ be mutually conjugate with respect to a symmetric positive definite matrix $A \in \mathbb{R}^{n \times n}$. Consider a linear combination which is equal to zero:

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i = 0$$

we pre-multiply by the matrix A

$$A \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i = \sum_{i=1}^{n} \alpha_i A \boldsymbol{u}_i = 0$$

and take the inner product with $\boldsymbol{u}_k$

$$\left(\boldsymbol{u}_k, \sum_{i=1}^{n} \alpha_i A \boldsymbol{u}_i \right) = \sum_{i=1}^{n} \alpha_i (\boldsymbol{u}_k, A\boldsymbol{u}_i) = \alpha_k (\boldsymbol{u}_k, A\boldsymbol{u}_k) = 0$$

Now, since A is positive definite, we have

$$(\boldsymbol{u}_k, A\boldsymbol{u}_k) > 0, \ \forall \boldsymbol{u}_k, \boldsymbol{u}_k \neq \mathbf{0}$$

Therefore, it must be that $\alpha_k = 0$, $\forall k$, which implies that $\boldsymbol{u}_i$, $i = 1, \ldots, n$ are linearly independent and since there are n of them, they form a basis for the $\mathbb{R}^n$ space.■

- What does it mean for a set of vectors to be linearly independent?

  Can you prove that a set of n linearly independent vectors in $\mathbb{R}^n$ form a basis for the $\mathbb{R}^n$ space?

Expansion of an Arbitrary Vector

Now consider an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^n$. We can expand $\boldsymbol{x}$ in our mutually conjugate basis as follows:

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i$$

where the scalar values α_i are to be determined. We next take the inner product of $\boldsymbol{u}_k$ with $A\boldsymbol{x}$:

$$(\boldsymbol{u}_k, A\boldsymbol{x}) = \left(\boldsymbol{u}_k, A\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i\right) = \left(\boldsymbol{u}_k, \sum_{i=1}^{n} \alpha_i A\boldsymbol{u}_i\right)$$

$$= \sum_{i=1}^{n} \alpha_i(\boldsymbol{u}_k, A\boldsymbol{u}_i) = \alpha_k(\boldsymbol{u}_k, A\boldsymbol{u}_k)$$

from which we can solve for the scalar coefficients as

$$\alpha_k = \frac{(\boldsymbol{u}_k, A\boldsymbol{x})}{(\boldsymbol{u}_k, A\boldsymbol{u}_k)}$$

and we have that an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^n$ can be expanded in terms of n mutually conjugate vectors $\boldsymbol{u}_i$, $i = 1, \ldots, n$ as

$$\boxed{\boldsymbol{x} = \sum_{i=1}^{n} \frac{(\boldsymbol{u}_k, A\boldsymbol{x})}{(\boldsymbol{u}_k, A\boldsymbol{u}_k)} \boldsymbol{u}_i}$$

Definition: If a minimization method always locates the minimum of a general quadratic function in no more than a predetermined number of steps directly related to number of variables n, then the method is called *quadratically convergent*.❐

Theorem: If a quadratic function $Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$ is minimized sequentially once along each direction of a set of n linearly independent, A-conjugate directions, then the global minimum of Q will be located at or before the n^{th} step regardless of the starting point.❐

<u>Proof</u>: We know that

$$\nabla Q(\boldsymbol{x}^*) = \boldsymbol{b} + A\boldsymbol{x}^* = \boldsymbol{0} \qquad \textbf{(1)}$$

and given $\boldsymbol{u}_i$, $i = 1, \ldots, n$ to be A-conjugate vectors or, in this case, directions of minimization, we know from previous theorem that they are linearly independent. Let $\boldsymbol{x}^1$ be the starting point of our search, then expanding the minimum $\boldsymbol{x}^*$ as

$$\boldsymbol{x}^* = \boldsymbol{x}^1 + \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i \qquad \textbf{(2)}$$

$$\begin{aligned}\boldsymbol{b} + A\boldsymbol{x}^* &= \boldsymbol{b} + A\left(\boldsymbol{x}^1 + \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i\right) \\ &= \boldsymbol{b} + A\boldsymbol{x}^1 + A\sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i = \boldsymbol{b} + A\boldsymbol{x}^1 + \sum_{i=1}^{n} \alpha_i A\boldsymbol{u}_i = \boldsymbol{0}\end{aligned}$$

taking the inner product with $\boldsymbol{u}_j$ (using the notation $\boldsymbol{v}^T\boldsymbol{u} = (\boldsymbol{v}, \boldsymbol{u})$) we have

$$\boldsymbol{u}_j^T(\boldsymbol{b} + A\boldsymbol{x}^1) + \boldsymbol{u}_j^T \sum_{i=1}^{n} \alpha_i A \boldsymbol{u}_i = \boldsymbol{u}_j^T(\boldsymbol{b} + A\boldsymbol{x}^1) + \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_j^T A \boldsymbol{u}_i = 0$$

which, since the $\boldsymbol{u}_i$ vectors are mutually conjugate with respect to the matrix A, we have

$$\boldsymbol{u}_j^T(\boldsymbol{b} + A\boldsymbol{x}^1) + \alpha_j \boldsymbol{u}_j^T A \boldsymbol{u}_j = 0$$

which can be re-written as

$$(\boldsymbol{b} + A\boldsymbol{x}^1)^T \boldsymbol{u}_j + \alpha_j \boldsymbol{u}_j^T A \boldsymbol{u}_j = 0.$$

Solving for the coefficients we have

$$\alpha_j = -\frac{(\boldsymbol{b} + A\boldsymbol{x}^1)^T \boldsymbol{u}_j}{\boldsymbol{u}_j^T A \boldsymbol{u}_j}. \tag{3}$$

Now in an iterative scheme where we determine successive approximations along the $\boldsymbol{u}_i$ directions by minimization, we have

$$\boldsymbol{x}^{i+1} = \boldsymbol{x}^i + \lambda_i^* \boldsymbol{u}_i,\ i = 1, \ldots, N \tag{4}$$

where the λ_i^* are found by minimizing $Q(\boldsymbol{x}^i + \lambda_i \boldsymbol{u}_i)$ with respect to the variable λ_i, and N is possibly greater than n.

Therefore, letting $\boldsymbol{y}^i = \boldsymbol{x}^{i+1} = \boldsymbol{x}^i + \lambda_i \boldsymbol{u}_i$, we set the derivative of $Q(\boldsymbol{y}^i(\lambda_i)) = Q(\boldsymbol{x}^i + \lambda_i \boldsymbol{u}_i)$ with respect to λ_i equal to 0 using the chain rule of differentiation:

$$\left.\frac{d}{d\lambda_i} Q(\boldsymbol{x}^{i+1})\right|_{\lambda_i^*} = \sum_{j=1}^{n} \frac{\partial Q}{\partial y_i^j}\left(\frac{\partial y_i^j}{\partial \lambda_i}\right) = \boldsymbol{u}_i^T \nabla Q(\boldsymbol{x}^{i+1}) = 0$$

but $\nabla Q(\boldsymbol{x}^{i+1}) = \boldsymbol{b} + A\boldsymbol{x}^{i+1}$ and therefore

$$\boldsymbol{u}_i^T(\boldsymbol{b} + A(\boldsymbol{x}^i + \lambda_i \boldsymbol{u}_i)) = 0$$

from which we get that the λ_i^* are given by

$$\lambda_i^* = -\frac{(\boldsymbol{b} + A\boldsymbol{x}^i)^T \boldsymbol{u}_i}{\boldsymbol{u}_i^T A \boldsymbol{u}_i} = -\frac{\boldsymbol{b}^T \boldsymbol{u}_i + \boldsymbol{x}^{iT} A \boldsymbol{u}_i}{\boldsymbol{u}_i^T A \boldsymbol{u}_i}. \quad \textbf{(5)}$$

From (4), we can write

$$\boldsymbol{x}^{i+1} = \boldsymbol{x}^i + \lambda_i^* \boldsymbol{u}_i = \boldsymbol{x}^1 + \sum_{j=1}^{i} \lambda_j^* \boldsymbol{u}_j$$

$$\boldsymbol{x}^i = \boldsymbol{x}^1 + \sum_{j=1}^{i-1} \lambda_j^* \boldsymbol{u}_j.$$

Forming the product $\boldsymbol{x}^{iT} A \boldsymbol{u}_i$ in (5) we get

$$\boldsymbol{x}^{iT} A \boldsymbol{u}_i = (\boldsymbol{x}^1)^T A \boldsymbol{u}_i + \sum_{j=1}^{i-1} \lambda_j^* \boldsymbol{u}_j^T A \boldsymbol{u}_i = (\boldsymbol{x}^1)^T A \boldsymbol{u}_i$$

because $\boldsymbol{u}_j^T A \boldsymbol{u}_i = 0$ for $j \neq i$. Therefore, the λ_i^* can be written as

$$\lambda_i^* = -\frac{(\boldsymbol{b} + A\boldsymbol{x}^1)^T \boldsymbol{u}_i}{\boldsymbol{u}_i^T A \boldsymbol{u}_i} \quad \textbf{(6)}$$

but comparing this (3) we see that $\lambda_i^* = \alpha_i$ and therefore

$$\boldsymbol{x}^* = \boldsymbol{x}^1 + \sum_{j=1}^{n} \lambda_j^* \boldsymbol{u}_j \quad \textbf{(7)}$$

which says that starting at $\boldsymbol{x}^1$ we take n steps of "length" λ_j^*, given by (6), in the $\boldsymbol{u}_j$ directions and we get the minimum.

Therefore $\boldsymbol{x}^*$ is reached in n steps or less if some $\lambda_j^* = 0$.■

Example: consider the quadratic function of two variables given as $f(\boldsymbol{x}) = 1 + x_1 - x_2 + x_1^2 + x_2^2$. Use the previous theorem to find the minimum starting at the origin and minimizing successively along the two directions given by the unit vectors $\boldsymbol{u}_1^T = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $\boldsymbol{u}_2^T = \begin{bmatrix} 0 & 1 \end{bmatrix}$. (First show that these vectors are mutually conjugate with respect to the Hessian matrix of the function.)

Solution: first write the function in matrix form as

$$f(\boldsymbol{x}) = 1 + \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = c + \boldsymbol{b}^T \boldsymbol{x} + \frac{1}{2} \boldsymbol{x}^T A \boldsymbol{x}$$

where we can clearly see the Hessian matrix A. We can now check that the two directions given are mutually conjugate with respect to A as

$$\boldsymbol{u}_1^T A \boldsymbol{u}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0, \; \boldsymbol{u}_1^T A \boldsymbol{u}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2,$$

$$\boldsymbol{u}_2^T A \boldsymbol{u}_2 = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4.$$

Starting from $\boldsymbol{x}^1 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ we find the two lengths, λ_1^* and λ_2^*, from (6) as

$$\lambda_1^* = -\frac{(\boldsymbol{b} + A\boldsymbol{x}^1)^T \boldsymbol{u}_1}{\boldsymbol{u}_1^T A \boldsymbol{u}_1} = -\frac{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{2} = -\frac{1}{2}$$

$$\lambda_2^* = -\frac{(\boldsymbol{b} + A\boldsymbol{x}^1)^T \boldsymbol{u}_2}{\boldsymbol{u}_2^T A \boldsymbol{u}_2} = -\frac{\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{4} = -\frac{1}{4}$$

and therefore, from (7), the minimum is found as

$$\boldsymbol{x}^* = \boldsymbol{x}^1 + \sum_{j=1}^{2} \lambda_j^* \boldsymbol{u}_j = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/4 \end{bmatrix}.$$

This can be checked by applying the formula $\boldsymbol{x}^* = -A^{-1}\boldsymbol{b}$.■

Note that the lengths λ_j^* calculated from (6) dependent only on the mutually conjugate directions themselves and the initial starting point, but not on the intermediate successive search points $\boldsymbol{x}^i$ with $i > 1$.

Thus, if we always start from the origin, then the minimum of a quadratic function can be written as

$$\boldsymbol{x}^* = -\sum_{i=1}^{n} \frac{\boldsymbol{b}^T \boldsymbol{u}_i}{\boldsymbol{u}_i^T A \boldsymbol{u}_i} \boldsymbol{u}_i. \qquad \textbf{(8)}$$

Of course, we still need a method of finding n A-conjugate vectors in $\subsetneq^n$ space.

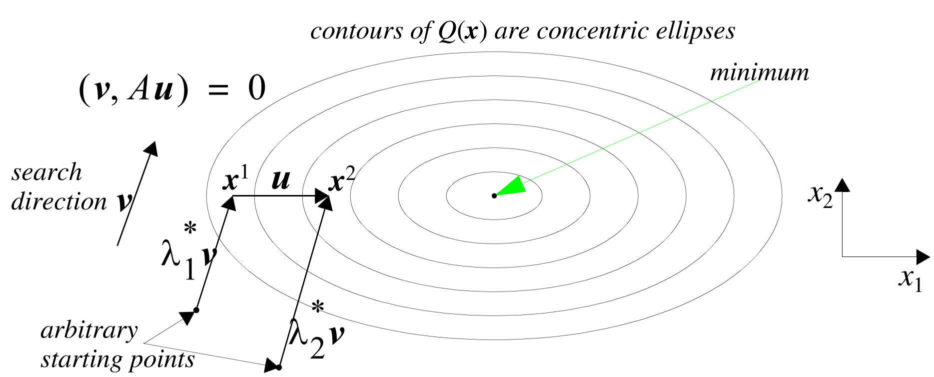
- The following theorem which we will not prove gives us a powerful technique for finding such minimization directions.

Theorem: Parallel Subspace Property

Given a direction $\boldsymbol{v}$ and a quadratic function $Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$, then starting from two different points, but arbitrary, we can determine the minimum in the $\boldsymbol{v}$ direction as $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$. The new direction $\boldsymbol{u} = \boldsymbol{x}^2 - \boldsymbol{x}^1$ is A-conjugate to $\boldsymbol{v}$, *i.e.* $(\boldsymbol{v}, A\boldsymbol{u}) = 0$.❐

Powell's Conjugate Direction Method

- The idea behind Powell's method is to use the parallel subspace property to create a set of conjugate directions.
- It then uses line searches along these "conjugate" directions to find the local minimum.
- Before we describe Powell's method it is instructive to consider the parallel subspace property geometrically in two dimensions as shown in the figure.
- The concentric ellipses are the contour lines of a quadratic function $Q(\boldsymbol{x})$ having a Hessian matrix A.
- Starting at the two arbitrary points shown we minimize along the $\boldsymbol{v}$ direction to arrive at points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$.
- The direction $\boldsymbol{u} = \boldsymbol{x}^2 - \boldsymbol{x}^1$ will be A-conjugate to $\boldsymbol{v}$.
- If we were to perform a further minimization along $\boldsymbol{u}$ it is clear that we would arrive at the minimum.



Graphical depiction of the parallel subspace concept used in Powell's method.

Powell's Method in Words

- In words, Powell's method to minimize a function $f(\boldsymbol{x})$ in $\mathbb{R}^n$ can be described as follows.
- First, initialize n search directions $\boldsymbol{s}_i$, $i = 1, \ldots n$ to the coordinate unit vectors $\boldsymbol{e}_i$, $i = 1, \ldots n$.
- Then, starting at an initial guess, $\boldsymbol{x}^0$, perform and initial search in the $\boldsymbol{s}_n$ direction which gets you to the point $\boldsymbol{X}$.
- Store $\boldsymbol{X}$ in $\boldsymbol{Y}$ and then update $\boldsymbol{X}$ by performing n successive minimizations along the n search directions.
- Create a new search direction, $\boldsymbol{s}_{n+1} = \boldsymbol{X} - \boldsymbol{Y}$ and minimize along this direction as well.
- After this last search we check for convergence by comparing the relative change in function value at the most recent $\boldsymbol{X}$ with respect to the function value at $\boldsymbol{Y}$.
- If we have not converged, then we discard the first search direction $\boldsymbol{s}_1$ and let $\boldsymbol{s}_i = \boldsymbol{s}_{i+1}$, $i = 1, \ldots n$ and repeat.

Algorithm: Powell's Method

input: $f(\boldsymbol{x})$, $\boldsymbol{x}^0$, ε, max_iteration

set: $\boldsymbol{s}_i = \boldsymbol{e}_i,\ i = 1, \ldots n$

find λ^* which minimizes $f(\boldsymbol{x}^0 + \lambda^* \boldsymbol{s}_n)$

set: $\boldsymbol{X} = \boldsymbol{x}^0 + \lambda^* \boldsymbol{s}_n$, $C = \text{False}$, $k = 0$

while $C \equiv \text{False}$ repeat

set: $\boldsymbol{Y} = \boldsymbol{X}$, $k = k + 1$

for $i = 1(1)n$

find λ^* which minimizes $f(\boldsymbol{X} + \lambda^* \boldsymbol{s}_i)$

set: $\boldsymbol{X} = \boldsymbol{X} + \lambda^* \boldsymbol{s}_i$

end

set: $\boldsymbol{s}_{i+1} = \boldsymbol{X} - \boldsymbol{Y}$

find λ^* which minimizes $f(\boldsymbol{X} + \lambda^* \boldsymbol{s}_{i+1})$

set: $\boldsymbol{X} = \boldsymbol{X} + \lambda^* \boldsymbol{s}_{i+1}$

if $k >$ max_iteration OR $|f(\boldsymbol{X}) - f(\boldsymbol{Y})| / \max\ [|f(\boldsymbol{X})|, 10^{-10}] < \varepsilon$

$C = \text{True}$

else

for $i = 1(1)n$

set: $\boldsymbol{s}_i = \boldsymbol{s}_{i+1}$

end

end

end

Example: Powell's Conjugate Direction Method

Consider the following function of two variables:

$$f(\boldsymbol{x}) = 2x_1^3 + x_1 x_2^3 - 10x_1 x_2 + x_2^2$$

starting at $\boldsymbol{x}^0 = \begin{bmatrix} 5 & 2 \end{bmatrix}^T$, $f(\boldsymbol{x}^0) = 314$ we perform one iteration of Powell's conjugate direction method.

Solution:

First we choose the n search directions as coordinate directions:

$$\boldsymbol{s}_1 = \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{s}_2 = \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and perform three successive searches starting at $\boldsymbol{Y} = \boldsymbol{X} = \boldsymbol{x}^0 = \begin{bmatrix} 5 & 2 \end{bmatrix}^T$ along $\boldsymbol{s}_2$, $\boldsymbol{s}_1$, and $\boldsymbol{s}_2$:

1. $$\min_{\lambda} f(\boldsymbol{X} + \lambda \boldsymbol{s}_2) = f\left(\begin{bmatrix} 5 \\ 2 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 250 + 5(2+\lambda)^3 - 50(2+\lambda) + (2+\lambda)^2 = F(\lambda)$$

$$\left.\frac{dF}{d\lambda}\right|_{\lambda^*} = 15(2+\lambda^*)^2 - 50 + (2+\lambda^*) = 0, \ 15(\lambda^*)^2 + 61\lambda^* + 12 = 0$$

$$\Rightarrow \lambda^* = \frac{-61 \pm \sqrt{3001}}{30} = \begin{cases} -0.20728721 \\ -3.8593795 \end{cases}, \begin{cases} F(-0.20728721) = 192.38545 \\ F(-3.8593795) = 314.28418 \end{cases}$$

$$\Rightarrow \lambda^* = -3.8593795 \quad \Rightarrow \boldsymbol{X} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} + \lambda^* \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -1.86 \end{bmatrix}$$

2. $\min_{\lambda} f(\boldsymbol{X}+\lambda \boldsymbol{s}_1) = f\left(\begin{bmatrix} 5 \\ -1.86 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = 2(5+\lambda)^3 + 12.165377(5+\lambda) + 3.457292$

$$\left.\frac{dF}{d\lambda}\right|_{\lambda^*} = 6(5+\lambda^*)^2 + 12.165377 = 0$$

$$\Rightarrow \lambda^* = \begin{cases} -3.5760748 \\ -6.4239252 \end{cases}, \begin{cases} F(-3.5760748) = 26.554075 \\ F(-6.4239252) = -19.639491 \end{cases}$$

$$\Rightarrow \lambda^* = -6.4239252 \quad \Rightarrow \boldsymbol{X} = \begin{bmatrix} 5 \\ -1.86 \end{bmatrix} + \lambda^* \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1.42 \\ -1.86 \end{bmatrix}$$

3. $\min_{\lambda} f(\boldsymbol{X}+\lambda \boldsymbol{s}_2) = f\left(\begin{bmatrix} -1.42 \\ -1.86 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$

$$= -5.726576 - 1.42(-1.86+\lambda)^3 + 14.2(-1.86+\lambda) + (-1.86+\lambda)^2$$

$$\left.\frac{dF}{d\lambda}\right|_{\lambda^*} = -4.26(-1.86+\lambda^*)^2 + 14.2 + 2(-1.86+\lambda^*) = 0$$

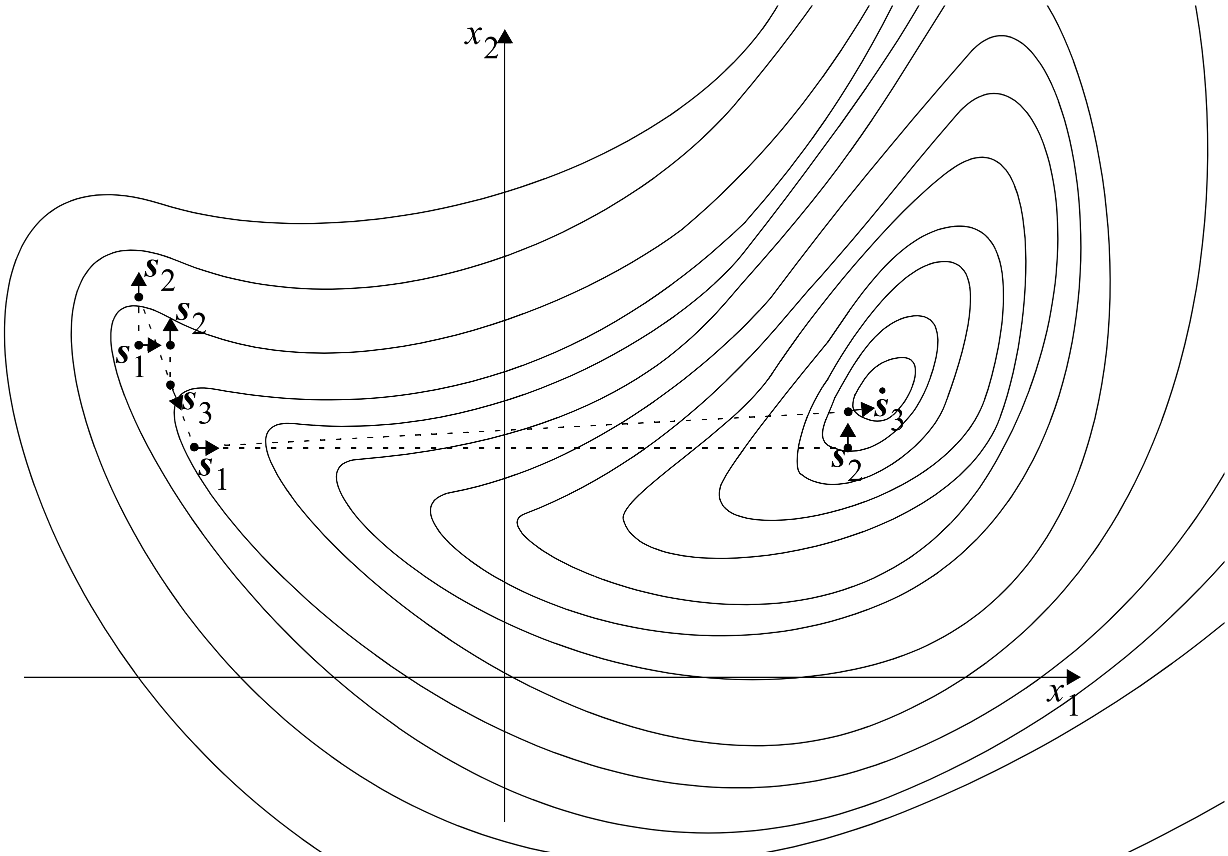
$$-4.26(\lambda^*)^2 + 17.8472\lambda^* - 4.257896 = 0$$

$$\Rightarrow \lambda^* = \frac{-17.8472 \pm 15.683367}{-8.52} = \begin{cases} 0.25397101 \\ 3.9355126 \end{cases},$$

$$\begin{cases} F(0.25397101) = -20.0 \\ F(3.9355126) = 15.357527 \end{cases}$$

$$\Rightarrow \lambda^* = 0.25397101 \quad \Rightarrow \boldsymbol{X} = \begin{bmatrix} -1.42 \\ -1.86 \end{bmatrix} + \lambda^* \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1.42 \\ -1.60 \end{bmatrix}$$

4. Now we set $\boldsymbol{s}_3 = \boldsymbol{X} - \boldsymbol{Y} = \begin{bmatrix} -1.42 \\ -1.60 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} -6.42 \\ -3.6 \end{bmatrix}$ and perform one more search in this direction before checking for convergence.



Geometrical view of Powell’s method after 2 iterations in the main loop.