# E(y) descent rate

How much closer does one step bring us to the solution **y* = 0**? In other words, how much smaller is ***e(y<sub>k+1</sub>)*** relatively to ***e(y<sub>k</sub>)***? From the definition of ***e(y)*** and for ***y<sub>k+1</sub>***, we obtain:

$$
\begin{aligned}
\frac{e(\mathbf{y}_k)-e(\mathbf{y}_{k+1})}{e(\mathbf{y}_k)} &= \frac{\mathbf{y}_k^T Q \mathbf{y}_k - \mathbf{y}_{k+1}^T Q \mathbf{y}_{k+1}}{\mathbf{y}_k^T Q \mathbf{y}_k} \\
&= \frac{\mathbf{y}_k^T Q \mathbf{y}_k - \left(\mathbf{y}_k - \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T Q \mathbf{g}_k}\mathbf{g}_k\right)^T Q \left(\mathbf{y}_k - \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T Q \mathbf{g}_k}\mathbf{g}_k\right)}{\mathbf{y}_k^T Q \mathbf{y}_k} \\
&= \frac{2\frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T Q \mathbf{g}_k}\mathbf{g}_k^T Q \mathbf{y}_k - \left(\frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T Q \mathbf{g}_k}\right)^2 \mathbf{g}_k^T Q \mathbf{g}_k}{\mathbf{y}_k^T Q \mathbf{y}_k} \\
&= \frac{2\mathbf{g}_k^T \mathbf{g}_k \mathbf{g}_k^T Q \mathbf{y}_k - (\mathbf{g}_k^T \mathbf{g}_k)^2}{\mathbf{y}_k^T Q \mathbf{y}_k \, \mathbf{g}_k^T Q \mathbf{g}_k}\,. \qquad (4)
\end{aligned}
$$

Since Q is invertible, we have:

$$\mathbf{g}_k = Q\mathbf{y}_k \quad \Rightarrow \quad \mathbf{y}_k = Q^{-1}\mathbf{g}_k$$

And

$$\mathbf{y}_k^T Q \mathbf{y}_k = \mathbf{g}_k^T Q^{-1} \mathbf{g}_k$$

,what allows to rewrite (4) as:

$$\frac{e(\mathbf{y}_k) - e(\mathbf{y}_{k+1})}{e(\mathbf{y}_k)} = \frac{(\mathbf{g}_k^T \mathbf{g}_k)^2}{\mathbf{g}_k^T Q^{-1} \mathbf{g}_k \, \mathbf{g}_k^T Q \mathbf{g}_k}$$

or

$$e(\mathbf{y}_{k+1}) = \left(1 - \frac{(\mathbf{g}_k^T \mathbf{g}_k)^2}{\mathbf{g}_k^T Q^{-1} \mathbf{g}_k \, \mathbf{g}_k^T Q \mathbf{g}_k}\right) e(\mathbf{y}_k)$$

# Kantorovich inequality.

**Kantorovich inequality**: Let Q be a positive definite, symmetric, n x n matrix.

Then, for any vector y there holds:

$$\frac{(\mathbf{y}^T\mathbf{y})^2}{\mathbf{y}^T Q^{-1}\mathbf{y}\,\mathbf{y}^T Q\mathbf{y}} \geq \frac{4\sigma_1\sigma_n}{(\sigma_1+\sigma_n)^2}$$

This inequality allows to prove the **Steepest Descent Rate** theorem.

# **Steepest Descent Rate** Theorem

**Steepest Descent Rate** theorem:

Let

$$f(\mathbf{x}) = c + \mathbf{a}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T Q \mathbf{x}$$

be a quadratic function of x, with Q symmetric and positive definite .

For any $x_o$, the method of steepest descent

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_k^T Q \mathbf{g}_k} \mathbf{g}_k$$

,where

$$\mathbf{g}_k = \mathbf{g}(\mathbf{x}_k) = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_k} = \mathbf{a} + Q\mathbf{x}_k$$

Converges to the unique minimum point

$$\mathbf{x}^* = -Q^{-1}\mathbf{a}$$

The difference at every step satisfies

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*) \leq \left(\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n}\right)^2 (f(\mathbf{x}_k) - f(\mathbf{x}^*))$$

, where $\sigma_1$ and $\sigma_n$ are the respectively the largest and the smallest singular values of Q

# Proof

From the definitions

$$\mathbf{y} = \mathbf{x} - \mathbf{x}^* \quad \text{and} \quad e(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T Q \mathbf{y}$$

we obtain

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*) \quad = \quad e(\mathbf{y}_{k+1}) = \left(1 - \frac{(\mathbf{g}_k^T \mathbf{g}_k)^2}{\mathbf{g}_k^T Q^{-1} \mathbf{g}_k \, \mathbf{g}_k^T Q \mathbf{g}_k}\right) e(\mathbf{y}_k) \leq$$

$$\leq \left(1 - \frac{4\sigma_1 \sigma_n}{(\sigma_1 + \sigma_n)^2}\right) e(\mathbf{y}_k) \quad = \quad \left(\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n}\right)^2 (f(\mathbf{x}_k) - f(\mathbf{x}^*))$$

Here, the Kantorovich inequality was used.

# Analysis

The ratio

$$\kappa(Q) = \sigma_1/\sigma_n$$

Is called the *condition number* of **Q**. The larger the condition number (ratio between the largest and the smallest singular values), the closer the ratio

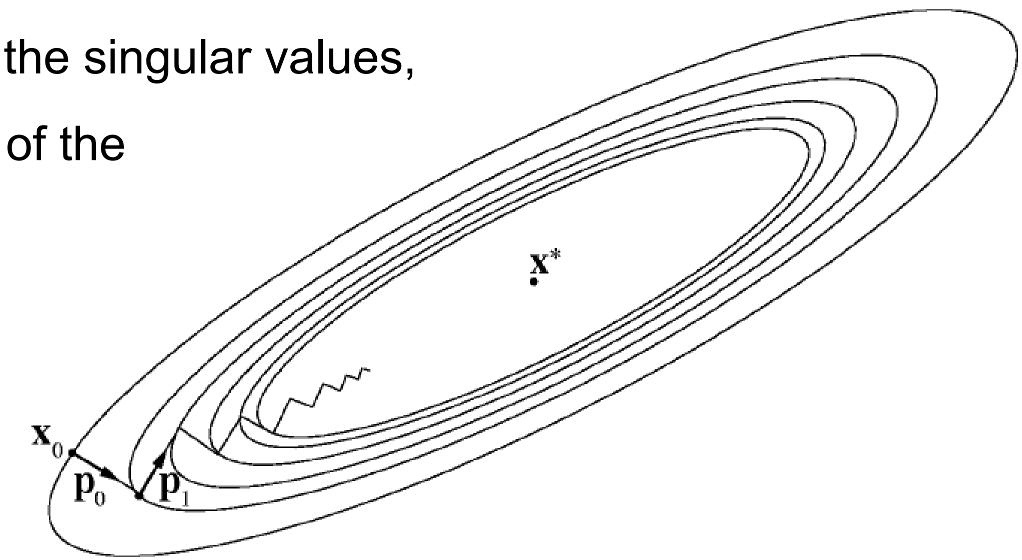
$$(\sigma_1 - \sigma_n)/(\sigma_1 + \sigma_n)$$

to 1 and therefore the slower the convergence.

# Illustration

Consider the two dimensional case, $\boldsymbol{x} \in \boldsymbol{R}^2$. The figure shows a trajectory $x_k$, imposed in the contours of f(x).

The greater the ratio between the singular values, of Q (which is the aspect ratio of the ellipses), the slower the convergence rate. If the isolines are circular (k(Q)=1) or the trajectory started from the ellipses axis, a single step brings us to **x***.



# Stopping criteria

We do not know the $\mathbf{x^*}$ and therefore $\mathbf{f(x^*)}$. Thus the stopping criteria is not trivial.

The criterion can be $|\mathbf{f(x_k)-f(x_{k-1})}|$ or $|\mathbf{x_k-x_{k-1}}|$. The second criterion is better, since it indicates proximity of $\mathbf{x^*}$.

# Line search

The steepest descent can be applied to general cases of **f**, not necessarily quadratic and not defined via

$$f(\mathbf{x}) = c + \mathbf{a}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T Q\mathbf{x} \qquad (1)$$

In these cases, **Q** is the matrix of the second derivatives of **f** with respect to **x**, called a *Hessian* of **f**. In this case, only n first derivatives are needed to calculate the direction $\mathbf{p}_k$. The step size requires calculation of Hessian of f(x), which requires computing $\binom{n}{2}$ second derivatives, and therefore is very expensive.

Using the line search allows to reach the minimum of f(x) in the direction $p_k$ without the Hessian calculation.