

Terminology

\exists means “there exists”

$\exists!$ means “there exists a unique”

\forall means “for all”

$:=$ means that the left-hand side is a new notation for the expression on the right-hand side

iff means “if and only if”

Conjugate directions

Motivation and definition

The conjugate gradient method is developed for a quadratic function of n variables

$$q(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{c}^T \mathbf{x}, \quad \text{with } \mathbf{H} \text{ positive definite.}$$

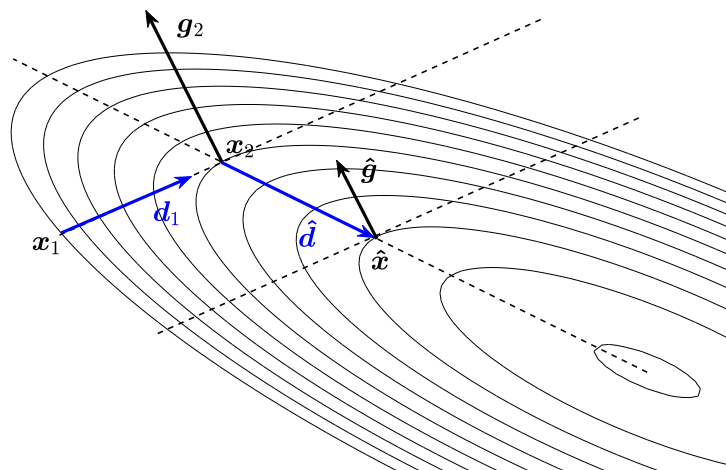
Let $\mathbf{g}_k := \nabla q(\mathbf{x}_k) = \mathbf{H} \mathbf{x}_k + \mathbf{c}$ and use exact line searches along directions \mathbf{d}_k (to be determined):

$$(EL) \quad \mathbf{g}_{k+1}^T \mathbf{d}_k = 0.$$

Important property:

$$(\Delta \mathbf{g}) \quad \begin{cases} \mathbf{g}_1 = \mathbf{H} \mathbf{x}_1 + \mathbf{c} \\ \mathbf{g}_2 = \mathbf{H} \mathbf{x}_2 + \mathbf{c} \end{cases} \Rightarrow \mathbf{g}_2 - \mathbf{g}_1 = \mathbf{H}(\mathbf{x}_2 - \mathbf{x}_1).$$

Idea of conjugate directions: (contours of a quadratic function)



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

Cartagena99

DEFINITION. The nonzero vectors $\mathbf{d}_1, \dots, \mathbf{d}_n$ are **H-conjugate** iff

$$\mathbf{d}_i^T \mathbf{H} \mathbf{d}_j = 0, \quad \forall i \neq j.$$

Note: \mathbf{H} positive definite $\Rightarrow \mathbf{d}_j^T \mathbf{H} \mathbf{d}_j > 0$.

LEMMA: The **H-conjugate** vectors $\mathbf{d}_1, \dots, \mathbf{d}_n$ are linearly independent.

PROOF: Given $\sum_{i=1}^n \lambda_i \mathbf{d}_i = \mathbf{0}$, multiply by $\mathbf{d}_j^T \mathbf{H}$ from the left with j arbitrary to get

$$\sum_{i=1}^n \lambda_i \mathbf{d}_j^T \mathbf{H} \mathbf{d}_i = 0 \Leftrightarrow \lambda_j \mathbf{d}_j^T \mathbf{H} \mathbf{d}_j = 0 \Leftrightarrow \lambda_j = 0.$$

□

Minimization along conjugate directions

With the minimizer at $\bar{\mathbf{x}}$, we can write (Exercise 1.5)

$$q(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{H}(\mathbf{x} - \bar{\mathbf{x}}) + a \quad \text{with} \quad a = q(\bar{\mathbf{x}}).$$

Assume that we have **H-conjugate** directions $\mathbf{d}_1, \dots, \mathbf{d}_n$ (how to find these comes later). Since these form a basis, we can make a variable change. With the invertible matrix

$$\mathbf{S} := (\mathbf{d}_1 \quad \mathbf{d}_2 \quad \dots \quad \mathbf{d}_n)$$

we can diagonalize \mathbf{H} :

$$\mathbf{S}^T \mathbf{H} \mathbf{S} = \begin{pmatrix} \mathbf{d}_1^T \\ \vdots \\ \mathbf{d}_n^T \end{pmatrix} \mathbf{H} (\mathbf{d}_1 \quad \mathbf{d}_2 \quad \dots \quad \mathbf{d}_n) = \begin{pmatrix} \mathbf{d}_1^T \mathbf{H} \mathbf{d}_1 & \dots & \mathbf{d}_1^T \mathbf{H} \mathbf{d}_n \\ \vdots & & \vdots \\ \mathbf{d}_n^T \mathbf{H} \mathbf{d}_1 & \dots & \mathbf{d}_n^T \mathbf{H} \mathbf{d}_n \end{pmatrix} = \text{diag}(\xi_1, \dots, \xi_n),$$

with $\xi_j := \mathbf{d}_j^T \mathbf{H} \mathbf{d}_j > 0$. Therefore, we make the coordinate change

$$\mathbf{x} = \tilde{\mathbf{x}}(\boldsymbol{\alpha}) \quad \text{with} \quad \tilde{\mathbf{x}}(\boldsymbol{\alpha}) = \mathbf{x}_1 + \mathbf{S} \boldsymbol{\alpha} = \mathbf{x}_1 + (\mathbf{d}_1 \quad \dots \quad \mathbf{d}_n) \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \mathbf{x}_1 + \sum_{j=1}^n \alpha_j \mathbf{d}_j.$$

The minimizer satisfies $\bar{\mathbf{x}} = \mathbf{x}_1 + \mathbf{S} \bar{\boldsymbol{\alpha}}$; hence, $\mathbf{x} - \bar{\mathbf{x}} = \mathbf{S}(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})$ and

$$(*) \quad q(\tilde{\mathbf{x}}(\boldsymbol{\alpha})) = \frac{1}{2}(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})^T \mathbf{S}^T \mathbf{H} \mathbf{S}(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) + a = \frac{1}{2} \sum_{j=1}^n \xi_j (\alpha_j - \bar{\alpha}_j)^2 + a.$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

occurs at $\lambda = \bar{\alpha}_1$, which defines $\mathbf{x}_2 := \mathbf{x}_1 + \bar{\alpha}_1 \mathbf{d}_1$. From this point, we search along \mathbf{d}_2 , which corresponds to setting $\boldsymbol{\alpha} = \boldsymbol{\alpha}_2 := (\bar{\alpha}_1, \lambda, 0, \dots, 0)^\top$, and the minimizer of

$$\begin{aligned} q(\mathbf{x}_2 + \lambda \mathbf{d}_2) &= q(\mathbf{x}_1 + \bar{\alpha}_1 \mathbf{d}_1 + \lambda \mathbf{d}_2) = q(\mathbf{x}_1 + \mathbf{S} \boldsymbol{\alpha}_2) \\ &= \frac{1}{2} \xi_2 (\lambda - \bar{\alpha}_2)^2 + \frac{1}{2} \sum_{i=3}^n \xi_i (0 - \bar{\alpha}_i)^2 + a, \end{aligned}$$

is clearly $\lambda = \bar{\alpha}_2$. Thus, we obtain

$$\begin{aligned} \mathbf{x}_2 &:= \mathbf{x}_1 + \bar{\alpha}_1 \mathbf{d}_1 = \mathbf{x}_1 + \mathbf{S} \bar{\boldsymbol{\alpha}}_1 = \tilde{\mathbf{x}}(\bar{\boldsymbol{\alpha}}_1) \quad \text{with} \quad \bar{\boldsymbol{\alpha}}_1 := (\bar{\alpha}_1 \ 0 \ \dots \ 0)^\top, \\ \mathbf{x}_3 &:= \mathbf{x}_2 + \bar{\alpha}_2 \mathbf{d}_2 = \mathbf{x}_1 + \mathbf{S} \bar{\boldsymbol{\alpha}}_2 = \tilde{\mathbf{x}}(\bar{\boldsymbol{\alpha}}_2) \quad \text{with} \quad \bar{\boldsymbol{\alpha}}_2 := (\bar{\alpha}_1 \ \bar{\alpha}_2 \ 0 \ \dots \ 0)^\top, \\ &\vdots \\ \mathbf{x}_{k+1} &:= \mathbf{x}_k + \bar{\alpha}_k \mathbf{d}_k = \mathbf{x}_1 + \mathbf{S} \bar{\boldsymbol{\alpha}}_k = \tilde{\mathbf{x}}(\bar{\boldsymbol{\alpha}}_k) \quad \text{with} \quad \bar{\boldsymbol{\alpha}}_k := (\bar{\alpha}_1 \ \dots \ \bar{\alpha}_k \ 0 \ \dots \ 0)^\top. \end{aligned}$$

(There is an explicit formula for $\bar{\alpha}_k$ for a quadratic function.) The minimizer is found after n line searches, since $q(\mathbf{x}_{n+1}) = q(\tilde{\mathbf{x}}(\bar{\boldsymbol{\alpha}}_n)) = a = q(\bar{\mathbf{x}})$. We have proved the following theorem.

THEOREM. *Minimization with exact line searches of a quadratic function along a sequence of \mathbf{H} -conjugate directions terminates after (at most) n steps.*

LEMMA CD. *Minimization with exact line searches of a quadratic function from the starting point \mathbf{x}_1 along a sequence of \mathbf{H} -conjugate directions $\mathbf{d}_1, \dots, \mathbf{d}_n$ results in a sequence of points $\mathbf{x}_2, \dots, \mathbf{x}_{n+1}$ such that, for $k = 1, \dots, n$,*

$$\mathbf{g}_{k+1}^\top \mathbf{d}_i = \nabla q(\mathbf{x}_{k+1})^\top \mathbf{d}_i = 0 \quad \text{for} \quad i = 1, \dots, k.$$

(The new gradient is perpendicular to all previous search directions.)

Remark: Since \mathbf{g}_{n+1} is perpendicular to all the basis vectors $\mathbf{d}_1, \dots, \mathbf{d}_n$, it has to be the zero vector.

PROOF: We differentiate (*), which is

$$q(\tilde{\mathbf{x}}(\boldsymbol{\alpha})) = \frac{1}{2} \sum_{i=1}^n \xi_i (\alpha_i - \bar{\alpha}_i)^2 + a, \quad \text{where} \quad \tilde{\mathbf{x}}(\boldsymbol{\alpha}) = \mathbf{x}_1 + \sum_{j=1}^n \alpha_j \mathbf{d}_j.$$

with respect to α_i and use the the chain rule for the left-hand side to get (∇ refers to derivatives with respect to \mathbf{x})

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

Cartagena99

Derivation of the Conjugate Gradient method

DEFINITION. A real inner product (scalar product) of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is a real-valued function denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$ with the properties

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} \rangle &= \langle \mathbf{v}, \mathbf{u} \rangle, \\ \langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle &= a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle, \quad a, b \in \mathbb{R}, \\ \langle \mathbf{u}, \mathbf{u} \rangle &\geq 0 \quad \text{with equality iff } \mathbf{u} = \mathbf{0}.\end{aligned}$$

One example is $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} = \mathbf{u}^T \mathbf{I} \mathbf{v}$.

Another is $\langle \mathbf{u}, \mathbf{v} \rangle_H = \mathbf{u}^T \mathbf{H} \mathbf{v}$ with \mathbf{H} positive definite.

We use the terminology

$$\mathbf{u} \text{ and } \mathbf{v} \text{ are } \mathbf{H}\text{-orthogonal} \Leftrightarrow \langle \mathbf{u}, \mathbf{v} \rangle_H = 0 \Leftrightarrow \mathbf{u} \text{ and } \mathbf{v} \text{ are } \mathbf{H}\text{-conjugate}$$

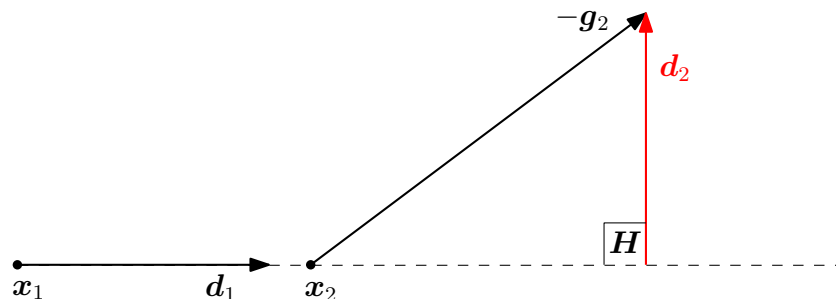
We now use the *Gram-Schmidt orthogonalization process* with $\langle \mathbf{u}, \mathbf{v} \rangle_H$ to derive the Conjugate Gradient method:

Step 0: Given the starting point \mathbf{x}_1 , compute $\mathbf{g}_1 = \nabla q(\mathbf{x}_1)$.

Step 1: Set $\mathbf{d}_1 := -\mathbf{g}_1$ (steepest descent)

Do exact line search, set $\mathbf{x}_2 := \mathbf{x}_1 + \bar{\alpha}_1 \mathbf{d}_1$ and compute $\mathbf{g}_2 = \nabla q(\mathbf{x}_2)$.

Step 2: Define \mathbf{d}_2 so that $\langle \mathbf{d}_2, \mathbf{d}_1 \rangle_H = 0$ using projection with respect to the \mathbf{H} -inner product:



Define (use the projection formula in linear algebra):

$$\mathbf{d}_2 := -\mathbf{g}_2 - \frac{\langle -\mathbf{g}_2, \mathbf{d}_1 \rangle_H}{\langle \mathbf{d}_1, \mathbf{d}_1 \rangle_H} \mathbf{d}_1$$

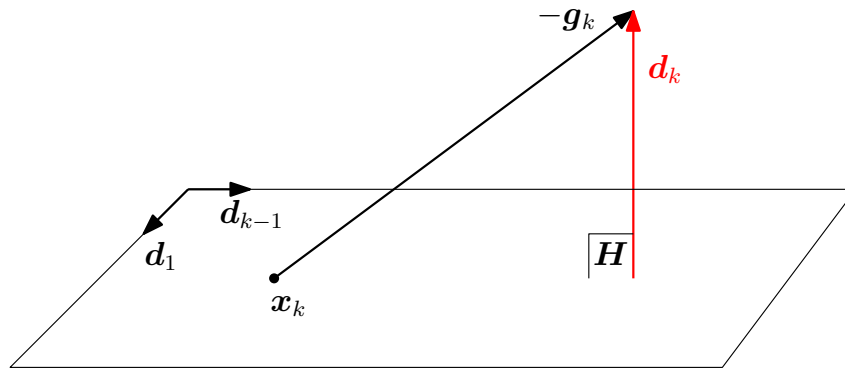
Do exact line search, set $\mathbf{x}_3 := \mathbf{x}_2 + \bar{\alpha}_2 \mathbf{d}_2$ and compute $\mathbf{g}_3 = \nabla q(\mathbf{x}_3)$.

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

Step k : Project $-\mathbf{g}_k$ onto the subspace spanned by $\mathbf{d}_1, \dots, \mathbf{d}_{k-1}$:



Define (sum of projections on each basis vector):

$$(D) \quad \begin{aligned} \mathbf{d}_k &:= -\mathbf{g}_k - \sum_{i=1}^{k-1} \frac{\langle -\mathbf{g}_k, \mathbf{d}_i \rangle_H}{\langle \mathbf{d}_i, \mathbf{d}_i \rangle_H} \mathbf{d}_i \\ &= -\mathbf{g}_k + \sum_{i=1}^{k-1} \beta_i \mathbf{d}_i, \quad \text{with } \beta_i := \frac{\langle \mathbf{g}_k, \mathbf{d}_i \rangle_H}{\langle \mathbf{d}_i, \mathbf{d}_i \rangle_H}. \end{aligned}$$

The task is now to simplify this formula. Recall:

$$(\Delta \mathbf{g}) \quad \mathbf{g}_2 - \mathbf{g}_1 = \mathbf{H}(\mathbf{x}_2 - \mathbf{x}_1)$$

$$(\text{Lemma CD}) \quad \mathbf{g}_{k+1}^\top \mathbf{d}_i = 0, \quad i = 1, \dots, k$$

We also need that (D) and Lemma CD imply

$$\mathbf{g}_{k+1}^\top \mathbf{g}_j = \mathbf{g}_{k+1}^\top \left(-\mathbf{d}_j + \sum_{i=1}^{j-1} \beta_i \mathbf{d}_i \right) = 0, \quad j = 1, \dots, k.$$

(The new gradient is also perpendicular to all previous gradients.) Numerator of β_i :

$$\begin{aligned} \langle \mathbf{g}_k, \mathbf{d}_i \rangle_H &= \mathbf{g}_k^\top \mathbf{H} \mathbf{d}_i = \frac{1}{\bar{\alpha}_i} \mathbf{g}_k^\top \mathbf{H}(\bar{\alpha}_i \mathbf{d}_i) = \frac{1}{\bar{\alpha}_i} \mathbf{g}_k^\top \mathbf{H}(\mathbf{x}_{i+1} - \mathbf{x}_i) \\ &\stackrel{(\Delta \mathbf{g})}{=} \frac{1}{\bar{\alpha}_i} \mathbf{g}_k^\top (\mathbf{g}_{i+1} - \mathbf{g}_i) = \begin{cases} 0 & \text{if } i < k-1, \\ \frac{1}{\bar{\alpha}_{k-1}} \mathbf{g}_k^\top \mathbf{g}_k & \text{if } i = k-1. \end{cases} \end{aligned}$$

Hence, (D) is simplified to $\mathbf{d}_k = -\mathbf{g}_k + \beta_{k-1} \mathbf{d}_{k-1}$. Denominator of β_i ($= \beta_{k-1}$):

$$\langle \mathbf{d}_i, \mathbf{d}_i \rangle_H = \mathbf{d}_i^\top \mathbf{H} \mathbf{d}_i = \frac{1}{\bar{\alpha}_i} \mathbf{d}_i^\top \mathbf{H}(\bar{\alpha}_i \mathbf{d}_i) = \frac{1}{\bar{\alpha}_i} \mathbf{d}_i^\top \mathbf{H}(\mathbf{x}_{i+1} - \mathbf{x}_i)$$

Cartagena99

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70

Conjugate gradient method for a general function $f(\mathbf{x})$:

Cyclic-coordinate search along the basis vectors $\mathbf{d}_1, \dots, \mathbf{d}_n$ defined by

$$\mathbf{d}_k = -\nabla f(\mathbf{x}_k) + \frac{\|\nabla f(\mathbf{x}_k)\|^2}{\|\nabla f(\mathbf{x}_{k-1})\|^2} \mathbf{d}_{k-1}$$

⊕ Only vectors involved – suitable for large problems

⊖ Requires accurate line searches

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, blue, serif font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue background with a white arrow pointing to the right, and a yellow arrow pointing to the left, both partially overlapping the text.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS
CALL OR WHATSAPP:689 45 44 70