The Conjugate Gradient Method

Luděk Kučera
MFF UK

25. května 2017

In the text, I describe how to use the method of conjugate gradients to solve and system of linear equations $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{b}$ is and given vector and $\mathbf{A}$ is symmetric and positive definite real matrix. The method is iterative: starting from an arbitrary point $\mathbf{x}_0$ in the corresponding Euklidean space, we will subsequently visit (by repeating always the same computation) points $\mathbf{x}_1, \mathbf{x}_2, \ldots$, until we eventually reach a point that is the solution of the system, or it is so close to the solution that the value will be sufficiently good approximate solution of the system of equations.

1 Solution of a system of linear equations by minimizing a quadratic functional

Let us define a functional f as follows:

$$f(x) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b} \qquad \text{for each } \mathbf{x} \in E_n.$$

Lemma 1 *The unique solution $\mathbf{x}_{sol}$ of the system $\mathbf{Ax} = \mathbf{b}$ is the unique minimum of the functional f.*

Důkaz: Consider a vector $\mathbf{x} \in E_n$ and the solution $\mathbf{x}_{sol}$ of the system $\mathbf{Ax} = \mathbf{b}$. Put $\mathbf{e} = \mathbf{x} - \mathbf{x}_{sol}$. Then

$$f(\mathbf{x}) = f(\mathbf{x}_{sol} + \mathbf{e}) = \frac{1}{2}(\mathbf{x}_{sol} + \mathbf{e})^T\mathbf{A}(\mathbf{x}_{sol} + \mathbf{e}) - (\mathbf{x}_{sol} + \mathbf{e})^T\mathbf{b} =$$

$$= \frac{1}{2}(\mathbf{x}_{sol}^T\mathbf{A}\mathbf{x}_{sol} + \mathbf{x}_{sol}^T\mathbf{A}\mathbf{e} + \mathbf{e}^T\mathbf{A}\mathbf{x}_{sol} + \mathbf{e}^T\mathbf{A}\mathbf{e}) - \mathbf{x}_{sol}^T\mathbf{b} - \mathbf{e}^T\mathbf{b} =$$

$$= (\frac{1}{2}\mathbf{x}_{sol}^T\mathbf{A}\mathbf{x}_{sol} - \mathbf{x}_{sol}^T\mathbf{b}) + \mathbf{e}^T\mathbf{A}\mathbf{e} + \mathbf{e}^T(\mathbf{A}\mathbf{x}_{sol} - \mathbf{b}) = f(\mathbf{x}_{sol}) + \mathbf{e}^T\mathbf{A}\mathbf{e},$$

because $\mathbf{x}_{sol}^T\mathbf{A}\mathbf{e} = \mathbf{e}^T\mathbf{A}\mathbf{x}_{sol}$ and therefore $\frac{1}{2}(\mathbf{x}_{sol}^T\mathbf{A}\mathbf{e} + \mathbf{e}^T\mathbf{A}\mathbf{x}_{sol}) = \mathbf{e}^T\mathbf{A}\mathbf{x}_{sol}$.

Since $\mathbf{A}$ is positive definite, the value of the expression $\mathbf{e}^T\mathbf{A}\mathbf{e}$ is always nonnegative and it is equal to 0 (the smallest possible value) if and only if $\mathbf{e} = \mathbf{0}$, which is when $\mathbf{x} = \mathbf{x}_{sol}$. ♣

2 Minimalization of the functional in a given direction

Now, let us assume that we are in a point $\mathbf{x}$ and we will start to travel in the direction given by a vector $\mathbf{p}$ (we may travel in te forward and the backward direction). This means that we may visit points $\mathbf{x}+\alpha\mathbf{p}$, where α is a real number. Our goal is to determine, where on this path the functional f is minimized. The value of the functionalu f on the path is

$$f(\mathbf{x}+\alpha\mathbf{p}) = \frac{1}{2}(\mathbf{x}+\alpha\mathbf{p})^T\mathbf{A}(\mathbf{x}+\alpha\mathbf{p}) - (\mathbf{x}+\alpha\mathbf{p})^T\mathbf{b} =$$
$$= \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \alpha\mathbf{p}^T\mathbf{A}\mathbf{x} + \frac{1}{2}\alpha^2\mathbf{p}^T\mathbf{A}\mathbf{p} - \mathbf{x}^T\mathbf{b} - \alpha\mathbf{p}^T b = U + \alpha V + \alpha^2 W,$$

where $U = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}$, $V = \mathbf{p}^T\mathbf{A}\mathbf{x} - \mathbf{p}^T\mathbf{b}$, $W = \frac{1}{2}\mathbf{p}^T\mathbf{A}\mathbf{p}$.

Gien a fixed x and p, U, V and W are constants and the expression $f(\mathbf{x}+\alpha\mathbf{p})$ is a quadratic function of α. It is known that this funcion has its minimum, when its derivative by α is equal to 0. The derivative is generally equal to $2W\alpha + V$ and hence the minimum is obtained for $\alpha = -\frac{V}{2W}$, in other words

$$\alpha = \frac{\mathbf{p}^T(\mathbf{b}-\mathbf{A}\mathbf{x})}{\mathbf{p}^T\mathbf{A}\mathbf{p}}.$$

The mimimum on the path $\mathbf{x}+\alpha\mathbf{p}$ is in the point

$$\mathbf{x} + \frac{\mathbf{p}^T(\mathbf{b}-\mathbf{A}\mathbf{x})}{\mathbf{p}^T\mathbf{A}\mathbf{p}}\mathbf{p} \tag{1}$$

3 The steepest descent direction

Now, let us imagine that we are in the point $\mathbf{x}$ and we want to travel in the direction of the steepest descent of the functional f. What is this direction?

Take a vector $\mathbf{e}$ of the length 1 and let us start to travel from $\mathbf{x}$ in this direction. Put $\varphi(\alpha) = f(\mathbf{x}+\alpha\mathbf{e})$. For fixed $\mathbf{e}$ the rate of the change of the functionalu f in the point $\mathbf{x}$ when traveling in the direction $\mathbf{e}$ is equal to the value of the derivative of the function φ podle α v bodě $\alpha = 0$. Let us calculate:

$$\varphi(\alpha) = f(\mathbf{x}+\alpha\mathbf{e}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \alpha\mathbf{e}^T\mathbf{A}\mathbf{x} + \alpha^2\mathbf{e}^T\mathbf{A}\mathbf{e} - \mathbf{x}^T\mathbf{b} - \alpha\mathbf{e}^T\mathbf{b}$$
$$\frac{\partial\varphi(\alpha)}{\partial\alpha} = \mathbf{e}^T\mathbf{A}\mathbf{x} + 2\alpha\mathbf{e}^T\mathbf{A}\mathbf{e} - \mathbf{e}^T\mathbf{b}$$

and therefore the value of the derivative for $\alpha = 0$ is

$$\left.\frac{\partial\varphi(\alpha)}{\partial\alpha}\right|_{\alpha=0} = \mathbf{e}^T\mathbf{A}\mathbf{x} - \mathbf{e}^T\mathbf{b} = -\mathbf{e}^T(\mathbf{b}-\mathbf{A}\mathbf{x}).$$

Now, let us try to find out, in which direction has the smallest (the most negative) value, i.e., when the scalar product $\mathbf{e}^T(\mathbf{b}-\mathbf{A}\mathbf{x})$ hes the largest value. It is clear that this happens when $\mathbf{e}$ goes in the direction $\mathbf{b}-\mathbf{A}\mathbf{x}$.

4 The method of the steepest descent

If we have to solve the equation $\mathbf{Ax} = \mathbf{b}$, it is sufficient to find the minimum of the functional $f(x) = \frac{1}{2}\mathbf{x}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{b}$, that can be looked for iteratively as follows: start in an arbitrary point that is denoted as $\mathbf{x}_0$. Often, we choose $\mathbf{x}_0 = \mathbf{0}$, unless there is a good reason to choose another point. Start from this point in the direction of the steepest descent of the functional f, which is the direction given by the vector $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$ (unless we arrived to the solution of $Ax = b$ - and the computation finishes - the vector has positive length). Move in this direction to get as low as possible (withrespect to f). Where to move is described above.

Using the procedure of the preceding paragraph, we move to a point that will be denoted as $\mathbf{x}_1$, and the procedure is repeated again and again. This means that if we are in $\mathbf{x}_i$, we put $\mathbf{r}_i = \mathbf{b} - \mathbf{Ax}_i$ (note that $\mathbf{r}_i$ also shows how far we are with the computation, how far $\mathbf{Ax}_i$ is from the vector $\mathbf{b}$; if the difference is very small, we may decide to halt). We move in the direction r_i, again to the point that minimizes the functional f on this path and we denot it as $\mathbf{x}_{i+1}$, etc. We continue until $\mathbf{r}_i$ (which is usually called i-th *residuum*, this is why we use "r") is not equal to 0 (the solution found) or at least sufficiently small (a good approximation solution found).

This method is called the *method of the steepest descent*.

5 The methoda of conjugated gradients

In 1952 Magnus Hestenes and Eduard Stiefel noticed that in the ellipsoidal geometry of positive definite matrices it is more convenient to proceed in a way that differs from the steepest descent and, perhaps surprisingly, behaves better: the method is now called *Conjugated Gradient Method* or CG or CGM.

It is well known that if $\mathbf{A}$ is a positive definite and symmetric real matrix, then the formula $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{Ay}$ has properties of scalar product, i.e., $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with the equation only if $\mathbf{x} = \mathbf{0}$, the operation is symmetric and $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle$. We will say that two vectors $\mathbf{x}$ and $\mathbf{y}$ are *conjugated*, if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ (which means that they are "orthogonal" with respect to the scalar product determined by the matrix $\mathbf{A}$).

The method of conjugated gradients proceeds in a very similar way as the steepest descent method, but we will try to make the directions, in which we move in different rounds of iteration, mutually conjugated. The method that is used is very similar to the well known Gram-Schmidt orthogonalization.

The method starts again in an arbitrary point $\mathbf{x}_0$ and it creates a sequance $\mathbf{x}_0, \mathbf{r}_0, \mathbf{p}_k, \mathbf{x}_1, \mathbf{r}_1, \mathbf{p}_1, \ldots$ in the following way:

Conjugated Gradient Method (unpolished)
Input data:
an integer n (dimension of the problem)
symmetric positive definite matrix $\mathbf{A}$ of the size $n \times n$,
a vector $\mathbf{b}$ of dimension n (the right hand side of the system),
a starting vector $\mathbf{x}_0$.

pro $k = 0, 1, 2, \ldots$ {

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k \; ; \quad \text{if } \mathbf{r}_k = \mathbf{0}, \text{ halt;} \tag{2}$$

$$\mathbf{p}_k = \mathbf{r}_k - \sum_{j=0}^{k-1} \frac{\mathbf{r}_k^T \mathbf{A} \mathbf{p}_j}{\mathbf{p}_j^T \mathbf{A} \mathbf{p}_j} \mathbf{p}_j \; ; \tag{3}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\mathbf{p}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} \mathbf{p}_k \; ; \tag{4}$$

}

The first step is the same as in the steepest descent method; we start in the direction $\mathbf{p}_0$ that is equal to $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$. However, if, after k iterations, we arrive to the point $\mathbf{x}_k$, we are not going to proceed exactly in the direction of the steepest descent given by the residuum $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$, but in the direction $\mathbf{p}_k$, that is $\mathbf{r}_k$ modified so that it is *conjugated* to all directions $\mathbf{p}_0, \ldots, \mathbf{p}_{k-1}$, in which we have moved previously.

The modified direction $\mathbf{p}_k$ ($\mathbf{p}$ as 'progress') is obtained from $\mathbf{r}_k$ similarly as in the Gramm-Schmidt ortogonalization by subtracting appropriate multiples of the previous directions $\mathbf{p}_0 = \mathbf{r}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{k-1}$ from $\mathbf{r}_k$ to get a vector that is conjugated to all $\mathbf{p}_0 = \mathbf{r}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{k-1}$, i.e., "orthogonal" in the sense of the scalar product definovaned by the matrix $\mathbf{A}$.

This means that for each $j < k$ we want to subtract from the vector $\mathbf{r}_k$ such a multiple α_j of the vector $\mathbf{p}_j$, that the result of the subtraction is conjugated with $\mathbf{p}_j$. This implies the following equation for α_j: $(\mathbf{r}_k - \alpha_j \mathbf{p}_j)^T \mathbf{A} \mathbf{p}_j = \mathbf{0}$, or, equivalently, $\mathbf{r}_k^T \mathbf{A} \mathbf{p}_j = \alpha_j \mathbf{p}_j^T \mathbf{A} \mathbf{p}_j$, which implies

$$\alpha_j = \frac{\mathbf{r}_k^T \mathbf{A} \mathbf{p}_j}{\mathbf{p}_j^T \mathbf{A} \mathbf{p}_j}.$$

This gives the formula (3) for $\mathbf{p}_k$, as used above.

If such a vector $\mathbf{p}_k$ is not a zero vector, we move from $\mathbf{x}_k$ in the direction given by the vector $\mathbf{p}_k$ to the point $\mathbf{x}_{k+1}$ that minimizes the functional f alon that line. In view of (1) we get the formula (4) for x_{k+1}, as used above.

If it happens that $\mathbf{r}_k$ is the zero vector, the computation halts, because in such a caseě $\mathbf{x}_k$ is the solution of the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ and it is not necessary to compute any more.

We can see one possible problem, namely that it will be $\mathbf{p}_k = 0$, because in the subsequent computation of $\mathbf{x}_{k+1}$ in the formula (4) we would divide by 0. However, it will be proved later that this could not happen.

Now, let us prove three simple lemmae that have several important consequences.

Lemma 2 *If $\mathbf{p}_k$ is defined, then it is a linear combinations of vectors $\mathbf{r}_0, \dots, \mathbf{r}_k$.*

Důkaz: The lemma is proved by induction by k. For $k = 0$ the statement holds, because $\mathbf{p}_0 = \mathbf{r}_0$. And if it is satisfied for $k-1$, then (3) immediately implies that it is valid for k as well. ♣

Lemma 3 *If $\mathbf{r}_{k+1}$ is defined, then*

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{A}\mathbf{p}_k, \quad \textit{where} \quad \alpha_k = \frac{\mathbf{p}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{A}\mathbf{p}_k}.$$

Důkaz: In view of (4), it is

$$\mathbf{x}_{k+1} - \mathbf{x}_k = \frac{\mathbf{p}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{A}\mathbf{p}_k} \mathbf{p}_k,$$

and therefore

$$\mathbf{r}_{k+1} - \mathbf{r}_k = (\mathbf{b} - \mathbf{A}\mathbf{x}_{k+1}) - (\mathbf{b} - \mathbf{A}\mathbf{x}_k) = -(\mathbf{A}\mathbf{x}_{k+1} - \mathbf{A}\mathbf{x}_k) = -\frac{\mathbf{p}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{A}\mathbf{p}_k} \mathbf{A}\mathbf{p}_k.$$

♣

Lemma 4 (Residua are mutually orthogonal)
Suppose that for some non-negative number k the following holds: $\mathbf{x}_j$ and $\mathbf{r}_j$ are defined and $\mathbf{r}_j \neq \mathbf{0}$ for all $j = 0, 1, \dots, k$, and, moreover, $\mathbf{r}_i^T \mathbf{r}_j = 0$ for all integers i and j such that $0 \le i < j \le k$.
Then $\mathbf{p}_k \neq \mathbf{0}$, which implies that $\mathbf{x}_{k+1}$ and $\mathbf{r}_{k+1}$ are well defined as well and $\mathbf{r}_{k+1}^T \mathbf{r}_j = 0$ for $j = 0, 1, \dots, k$.

Důkaz: Since the rezidua $\mathbf{r}_0, \dots, \mathbf{r}_k$ are non-zero and mutually orthogonal, they are linearly independent. The sum in the formula (3) is a linear combination of vectors $\mathbf{p}_0, \dots, \mathbf{p}_{k-1}$ and therefore, in view of the preceding lemmae, they are also linear combinations of vectors $\mathbf{r}_0, \dots, \mathbf{r}_{k-1}$. In view of (3), p_k is a *non-trivial* linear combination of vectors $\mathbf{r}_0, \dots, \mathbf{r}_k$. This is why $\mathbf{p}_k$ can not be a zero vector. And if $\mathbf{p}_k \neq \mathbf{0}$, then the vectors $\mathbf{x}_{k+1}$ and $\mathbf{r}_{k+1}$ are well defined as well.

Finally, let us compute the value $\mathbf{r}_{k+1}^T \mathbf{r}_j$ for a given integer j in the range $0 \le j \le k$:

$$\mathbf{r}_{k+1}^T \mathbf{r}_j = (\mathbf{r}_k - \alpha_k \mathbf{A}\mathbf{p}_k)^T \mathbf{r}_j = \mathbf{r}_k^T \mathbf{r}_j - \alpha_k \mathbf{p}_k^T \mathbf{A} \left(\mathbf{p}_j + \sum_{i=0}^{j-1} \frac{\mathbf{r}_j^T \mathbf{A}\mathbf{p}_i}{\mathbf{p}_i^T \mathbf{A}\mathbf{p}_i} \mathbf{p}_i \right) = \mathbf{r}_k^T \mathbf{r}_j - \alpha_k \mathbf{p}_k^T \mathbf{A}\mathbf{p}_j,$$

where

$$\alpha_k = \frac{\mathbf{p}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k},$$

because $\mathbf{p}_k$ has been selected so that the following holds: $\mathbf{p}_k^T \mathbf{A} \mathbf{p}_i = 0$ for $i < k$.

Now, we know that for $j < k$ it is

$$\mathbf{p}_j^T \mathbf{r}_k = r_j - \sum_{i=0}^{j-1} \frac{\mathbf{r}_j^T \mathbf{A} \mathbf{p}_i}{\mathbf{p}_i^T \mathbf{A} \mathbf{p}_i} \mathbf{p}_i = \mathbf{r}_j \mathbf{r}_k,$$

and

$$\mathbf{p}_k^T \mathbf{r}_k = (\mathbf{r}_k - \sum_{j=0}^{k-1} \frac{\mathbf{r}_k^T \mathbf{A} \mathbf{p}_j}{\mathbf{p}_j^T \mathbf{A} \mathbf{p}_j} \mathbf{p}_j)^T \mathbf{r}_k = \mathbf{r}_k^T \mathbf{r}_k.$$

If now we have $j < k$, then the assumptions of the theorem imply that $\mathbf{r}_k^T \mathbf{r}_j = 0$ and $\mathbf{p}_k^T \mathbf{A} \mathbf{p}_j = 0$, and therefore $\mathbf{r}_{k+1}^T \mathbf{r}_j = 0$. If, conversely, $j = k$, then

$$\mathbf{r}_k^T \mathbf{r}_j - \alpha_k \mathbf{p}_k^T \mathbf{A} \mathbf{p}_j = \mathbf{r}_k^T \mathbf{r}_k - \mathbf{p}_k^T \mathbf{r}^k = \mathbf{r}_k^T \mathbf{r}_k - \mathbf{r}_k^T \mathbf{r}^k = 0.$$

♣

The lemma says that the residua are mutually orthogonal in the classical sense (the standard scalar product), but not conjugated in general (the scalar product induced by the matrix $\mathbf{A}$).

One of the very important consequence of the lemma is the following theorem (let us note that the theorem is valid under assumption of exact arithmetic, without rounding errors):

Věta 1 *The conjugated gradient method finds the exact solution of the system* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *of linear equations after at most* n *steps.*

Důkaz: Assume that the conjungated gradient method does not halt after n krocích. In such a case it would create non-zero vector $\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_n$ that, according to the previous lemma, are mutually orthogonal, and hence lineárly independent, which is not possible in a space of the dimension n. ♣

Anothe very useful lemma says that in the formula (3) for computing of $\mathbf{p}_k$, all terms in the sum, with the exception of the term corresponding to $j = k-1$, are equal to 0, which makes it possible to simplify greatly the formula as follows:

Lemma 5

$$\mathbf{r}_k^T \mathbf{A} \mathbf{p}_j = 0 \qquad \textit{for} \quad j = 0, \ldots, k-2$$

$$\frac{\mathbf{r}_{k+1}^T \mathbf{A} \mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k} = \frac{\mathbf{r}_{k+1}^T \mathbf{r}_{k+1}}{\mathbf{r}_k^T \mathbf{r}_k}$$

Důkaz: In view of the lemma 3**0** we know that

$$\mathbf{r}_k^T \mathbf{A} \mathbf{p}_j = \mathbf{r}_k^T (-\frac{1}{\alpha_{j-1}} (\mathbf{r}_j - \mathbf{r}_{j-1}) = 0 \qquad \text{for } j = 0, 1, \ldots, k-2$$

$$\frac{\mathbf{r}_{k+1}^T \mathbf{A}\mathbf{p}_k}{\mathbf{p}_k^T \mathbf{A}\mathbf{p}_k} = \frac{\mathbf{r}_{k+1}^T \mathbf{A}\mathbf{p}_{k-1}}{\alpha_{k-1}\mathbf{p}_{k-1}^T \mathbf{A}\mathbf{p}_{k-1}} = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}}$$

♣

Důsledek 1

$$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k, \qquad \textit{where } \beta_k = \frac{\mathbf{r}_{k+1}^T \mathbf{r}_{k+1}}{\mathbf{r}_k^T \mathbf{r}_k}$$

Důkaz: ♣

Conjugated Gradient Method (with indices)
Input data:
an integer n (dimension)
a symmetric positive definite matrix $\mathbf{A}$ of size $n \times n$,
a vector $\mathbf{b}$ of dimension n (the right hand side of the system),
a starting vector $\mathbf{x}_0$.

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$;
$\mathbf{p}_0 = \mathbf{r}_0$;
for $k = 0, 1, 2, \dots$ {

$$\begin{aligned}
\mathbf{z}_k &= \mathbf{A}\mathbf{p}_k \ ; \\
\alpha_k &= \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{p}_k^T \mathbf{z}_k} \ ; \\
\mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{p}_k \ ; \\
\mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k \mathbf{z}_k; \quad \text{if } \mathbf{r}_k = \mathbf{0}, \text{ halt;} \\
\beta_k &= \frac{\mathbf{r}_{k+1}^T \mathbf{r}_{k+1}}{\mathbf{r}_k^T \mathbf{r}_k} \ ; \\
\mathbf{p}_{k+1} &= \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k \ ;
\end{aligned}$$

}

Conjugated Gradient Method (final form)
Input data:
an integer n (dimension)
a symmetric positive definite matrix $\mathbf{A}$ of size $n \times n$,
a vector $\mathbf{b}$ of dimension n (the right hand side of the system),
a starting vector $\mathbf{x}_0$.

$\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}_0$;
$lrr = \mathbf{r}^T\mathbf{r}$;
$\mathbf{p} = \mathbf{r}$;
for $k = 0, 1, 2, \ldots$ {

$$
\begin{aligned}
\mathbf{z} &= \mathbf{A}\mathbf{p} \;; \\
\alpha &= \frac{lrr}{\mathbf{p}^T\mathbf{z}} \;; \\
\mathbf{x} &= \mathbf{x} + \alpha\mathbf{p} \;; \\
\mathbf{r} &= \mathbf{r} - \alpha\mathbf{z}; \quad \text{if } \mathbf{r} = \mathbf{0}\text{, halt;} \\
rr &= \mathbf{r}^T\mathbf{r} \;; \\
\beta &= \frac{rr}{lrr} \;; \\
lrr &= rr \\
\mathbf{p} &= \mathbf{r} + \beta\mathbf{p} \;;
\end{aligned}
$$

}