MATH 5330: Computational Methods of Linear Algebra

Lecture Note 4: The Conjugate Gradient Method

Xianyi Zeng
Department of Mathematical Sciences, UTEP

1 A Starting Point: The Steepest Descent Method

If $A \in \mathbb{R}^{n\times n}$ is symmetric positive-definite, solving:

$$A\boldsymbol{x} = \boldsymbol{b}\,, \tag{1.1}$$

is equivalent to minimizing the function:

$$\phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^t A\boldsymbol{x} - \boldsymbol{x}^t\boldsymbol{b}\,. \tag{1.2}$$

Thus we can apply any optimization algorithm to solve this minimization problem and obtain a method for solving (1.1).

At the point, let us consider the *steepest descent method* and select any initial guess $\boldsymbol{x}_0$. With $\boldsymbol{x}_k$ available we try to find the direction along which $\phi(\boldsymbol{x})$ decreases most rapidly starting from $\boldsymbol{x}_k$ and compute the next point $\boldsymbol{x}_{k+1}$ by minimizing $\phi(\boldsymbol{x})$ in this direction. By Taylor series expansion

$$\phi(\boldsymbol{x}_k + \alpha\boldsymbol{d}) = \phi(\boldsymbol{x}_k) + \alpha\nabla\phi(\boldsymbol{x}_k)^t\boldsymbol{d} + O(||\boldsymbol{d}||^2)\,;$$

thus the direction we're looking for is given by $-\nabla\phi(\boldsymbol{x}_k) = \boldsymbol{r}_k$, the *residual* at the k-th iteration:

$$\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k\,. \tag{1.3}$$

If $\boldsymbol{r}_k \neq 0$, we try to find the next solution point $\boldsymbol{x}_{k+1}$ by minimizing $\phi(\boldsymbol{x}_k + \alpha\boldsymbol{r}_k)$ for all $\alpha \in \mathbb{R}$ (called the *exact line search*). Note that $\phi(\boldsymbol{x}_k + \alpha\boldsymbol{r}_k)$ is a second-degree polynomial in α:

$$\phi(\boldsymbol{x}_k + \alpha\boldsymbol{r}_k) = \frac{1}{2}(\boldsymbol{r}_k^t A\boldsymbol{r}_k)\alpha^2 + (\boldsymbol{r}_k^t A\boldsymbol{x}_k - \boldsymbol{r}_k^t\boldsymbol{b})\alpha + \phi(\boldsymbol{x}_k)\,,$$

the solution to the exact line search is:

$$\alpha_k = \frac{\boldsymbol{r}_k^t\boldsymbol{r}_k}{\boldsymbol{r}_k^t A\boldsymbol{r}_k}\,;$$

the denominator is never zero due to the positive-definiteness of A and $\boldsymbol{r}_k \neq 0$, thus we can compute $\boldsymbol{x}_{k+1}$ as well as $\boldsymbol{r}_{k+1}$. This leads to the following (idealized) algorithm:

$$\begin{aligned}
&\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0\,, \quad k = 0\,; \\
&\text{while } \boldsymbol{r}_k \neq 0\,: \\
&\qquad \alpha_k = (\boldsymbol{r}_k^t\boldsymbol{r}_k)/(\boldsymbol{r}_k^t A\boldsymbol{r}_k)\,; \\
&\qquad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\boldsymbol{r}_k\,; \\
&\qquad \boldsymbol{r}_{k+1} = \boldsymbol{b} - A\boldsymbol{x}_{k+1}\,; \\
&\qquad k = k+1\,; \\
&\text{end} \\
&\boldsymbol{x} = \boldsymbol{x}_k\,.
\end{aligned} \tag{1.4}$$

A theoretical bound on the convergence of the steepest descent method is given by:

$$\left(\phi(\boldsymbol{x}_{k+1})+\frac{1}{2}\boldsymbol{b}^t A^{-1}\boldsymbol{b}\right)\leq\left(1-\frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}\right)\left(\phi(\boldsymbol{x}_k)+\frac{1}{2}\boldsymbol{b}^t A^{-1}\boldsymbol{b}\right), \tag{1.5}$$

where $-(\boldsymbol{b}^t A^{-1}\boldsymbol{b})/2$ is the theoretical minimum of $\phi(\boldsymbol{x})$, and $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ are the smallest and largest eigenvalues of A. Thus the global convergence is guaranteed due to the fact that $0<\lambda_{\min}(A)<\lambda_{\max}(A)$.

2 The Projection Method

An issue with the steepest descend method is that the decay rate of (1.5) can be extremely close to 1 in many practical applications, which means a large number of iterations is needed to achieve a certain accuracy.

The *projection method* can be viewed as an extension of the algorithm (1.4) in the following sense. In every iteration of the steepest descend algorithm, we compute a solution $\boldsymbol{x}_{k+1}$ in the one-dimensional affine space $\boldsymbol{x}_k+\text{span}(\boldsymbol{r}_k)$; thus it is also contained in the *affine space*:

$$\boldsymbol{x}_0+\text{span}(\boldsymbol{r}_0\,,\,\boldsymbol{r}_1\,,\,\cdots,\,\boldsymbol{r}_k)\,,$$

which has the dimension no larger than $\min(n,k+1)$.

Remark 1. *An affine space $\mathcal{A}\subseteq\mathcal{V}$, where $\mathcal{V}$ is a vector space, is not necessarily a vector space itself, at least in the sense that it does not necessarily contain a zero vector. A formal definition for the affine space requires that if $\boldsymbol{v}_1\,,\,\boldsymbol{v}_2\in\mathcal{A}$ and $\alpha\in\mathbb{R}$, then $(1-\alpha)\boldsymbol{v}_1+\alpha\boldsymbol{v}_2\in\mathcal{A}$. It is not difficult to show that let $\boldsymbol{v}_0\in\mathcal{A}$ be arbitrary, then the set:*

$$\{\boldsymbol{v}\,:\,\boldsymbol{v}+\boldsymbol{v}_0\in\mathcal{A}\}$$

is a linear subspace of $\mathcal{V}$. Denote this linear subspace by $\mathcal{S}$, then we write $\boldsymbol{A}=\boldsymbol{v}_0+\mathcal{S}$. One can check that $\mathcal{S}$ is independent of the particular choice of $\boldsymbol{v}_0$.

Remark 2. *Let $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k\in\mathcal{V}$, then* $\text{span}(\boldsymbol{v}_1\,,\,\boldsymbol{v}_1\,,\,\cdots,\,\boldsymbol{v}_k)$ *is the linear vector subspace that is composed of all linear combinations of $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k$; its dimension is at most* $\min(k,n)$.

The idea of the general projection method is to search for a "best" approximation in the affine space $\mathcal{V}_k=\boldsymbol{x}_0+\mathcal{K}_k$, where $\mathcal{K}_k$ is a linear subspace spanned by $k+1$ vectors $\boldsymbol{v}_0\,,\,\boldsymbol{v}_1\,,\,\cdots,\,\boldsymbol{v}_k$:

$$\mathcal{K}_k=\text{span}(\boldsymbol{v}_0\,,\,\boldsymbol{v}_1\,,\,\cdots,\,\boldsymbol{v}_k)\,. \tag{2.1}$$

In order to define the "best" approximation, a general way is to require that the residual $\boldsymbol{r}_k$ is orthogonal to another linear space $\mathcal{L}_k$ spanned by a set of $k+1$ vectors $\boldsymbol{w}_0\,,\,\boldsymbol{w}_1\,,\,\cdots,\,\boldsymbol{w}_k$. Note that if $\mathcal{L}_k=A\mathcal{K}_k$ this requirement is equivalent to the minimization problem:

$$\boldsymbol{x}_{k+1}=\arg\min_{\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}_k}||\boldsymbol{b}-A\boldsymbol{x}||_2\,.$$

Such projection methods thusly involve the following components:

- How to construct the vectors $\boldsymbol{v}_k$ and $\boldsymbol{w}_k$.
- How to solve the minimization problem efficiently in the k-th iteration.

A more extensive discussion will be provided in the next lecture on Krylov space methods. For now let us consider an alternative to define the "best" approximation as finding the minimum of $\phi(\boldsymbol{x})$ on the affine space $\boldsymbol{x}_0+\mathcal{K}_k$.

3 The Conjugate Gradient Method

Suppose the search spaces $\mathcal{K}_k$ are spanned by some linearly-independent vectors $\boldsymbol{p}_0$, $\cdots$, $\boldsymbol{p}_k$, we denote $P_k \in \mathbb{R}^{n\times(k+1)}$ as $P_k = [\boldsymbol{p}_0\ \boldsymbol{p}_1\ \cdots\ \boldsymbol{p}_k]$. Hence we may write $\boldsymbol{x}_{k+1} = \boldsymbol{x}_0 + P_k \boldsymbol{y}_k$ for some coefficient vector $\boldsymbol{y}_k \in \mathbb{R}^{k+1}$. For convenience we also write $\mathcal{K}_{-1} = \varnothing$ so that $\boldsymbol{x}_0$ minimize $\phi(\boldsymbol{x})$ in the affine space $\boldsymbol{x}_0 + \mathcal{K}_{-1} = \{\boldsymbol{x}_0\}$.

Now let $\boldsymbol{x}_k$ and $\boldsymbol{p}_0,\cdots,\boldsymbol{p}_k$ have already been computed, we consider minimizing $\phi(\boldsymbol{x})$ over the space $\boldsymbol{x}_0 + \mathcal{K}_k$ and write for now the solution as:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_0 + P_{k-1}\boldsymbol{y} + \alpha \boldsymbol{p}_k \,, \tag{3.1}$$

for some $\boldsymbol{y} \in \mathbb{R}^k$ and $\alpha \in \mathbb{R}$. The hope is that $\boldsymbol{y} = \boldsymbol{y}_{k-1}$ so that the minimization problem is equivalent to the line search along the direction $\boldsymbol{p}_k$, an easy sub-problem to solve. To this end we assume (3.1) and obtain

$$\phi(\boldsymbol{x}_{k+1}) = \phi(\boldsymbol{x}_0 + P_{k-1}\boldsymbol{y}) + \alpha \boldsymbol{y}^t P_{k-1}^t A \boldsymbol{p}_k + \frac{\alpha^2}{2} \boldsymbol{p}_k^t A \boldsymbol{p}_k - \alpha \boldsymbol{p}_k^t \boldsymbol{r}_0 \,.$$

If $\boldsymbol{p}_k$ is chosen such that $P_{k-1}^t A \boldsymbol{p}_k = 0$, i.e., $\boldsymbol{p}_j^t A \boldsymbol{p}_k = 0$ for all $j = 0,\cdots,k-1$, the second term vanishes and the contribution of $\boldsymbol{y}$ is contained in the first term. An equivalent statement of this condition is that $\boldsymbol{p}_k$ is *A-conjugate* to the subspace $\mathcal{K}_{k-1}$.

Assuming this condition is satisfied, we thusly have:

$$\phi(\boldsymbol{x}_{k+1}) = \phi(\boldsymbol{x}_0 + P_{k-1}\boldsymbol{y}) + \frac{\alpha^2}{2} \boldsymbol{p}_k^t A \boldsymbol{p}_k - \alpha \boldsymbol{p}_k^t \boldsymbol{r}_0 \,,$$

where the minimization over $(\boldsymbol{y},\alpha)$ is decoupled into minimizing over $\boldsymbol{y}$ and over α separately. The first part leads to $\boldsymbol{y} = \boldsymbol{y}_{k-1}$ as desired; and the second part leads to:

$$\alpha_k = \frac{\boldsymbol{p}_k^t \boldsymbol{r}_0}{\boldsymbol{p}_k^t A \boldsymbol{p}_k} = \frac{\boldsymbol{p}_k^t \boldsymbol{r}_k}{\boldsymbol{p}_k^t A \boldsymbol{p}_k} \,. \tag{3.2}$$

Here the second equality comes from:

$$\boldsymbol{p}_k^t \boldsymbol{r}_k = \boldsymbol{p}_k^t (\boldsymbol{b} - A(\boldsymbol{x}_0 + P_{k-1}\boldsymbol{y}_{k-1})) = \boldsymbol{p}_k^t \boldsymbol{r}_0 - (\boldsymbol{p}_k^t A P_{k-1})\boldsymbol{y}_{k-1} = \boldsymbol{p}_k^t \boldsymbol{r}_0 \,; \tag{3.3}$$

and one can compute the next iteration as $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$.

Now the problem reduces to finding $\boldsymbol{p}_k$ such that it is A-conjugate to $\mathcal{K}_{k-1}$. Let us denote this requirement as $\boldsymbol{p}_k \in (A\mathcal{K}_{k-1})^\perp$, where $A\mathcal{K}_{k-1}$ is the linear space obtained by pre-multiplying any member of $\mathcal{K}_{k-1}$ by the matrix A and the superscript $\perp$ denotes the orthogonal (w.r.t. the usual Euclidean norm) complement. First, we want to know whenever the current residual $\boldsymbol{r}_k$ is non-zero, the next non-zero search direction $\boldsymbol{p}_k \in (A\mathcal{K}_{k-1})^\perp$ so that $\boldsymbol{p}_k^t \boldsymbol{r}_k \neq 0$ (hence $\alpha_k \neq 0$) can always be found.

To this end, let us use the method of induction and first assume $\boldsymbol{r}_0 \neq 0$. Note that $\mathcal{K}_{-1} = \varnothing$ thus $(A\mathcal{K}_{-1})^\perp = \mathbb{R}^n$, and we can simply set $\boldsymbol{p}_0 = \boldsymbol{r}_0 \neq 0$. Now suppose we work up to the point that $\boldsymbol{r}_k \neq 0$ and want to find a suitable $\boldsymbol{p}_k$. Note that $\boldsymbol{x} = A^{-1}\boldsymbol{b}$, the exact solution, does not belong to $\boldsymbol{x}_0 + \mathcal{K}_{k-1}$ since $\boldsymbol{r}_k \neq 0$, we have:

$$A^{-1}\boldsymbol{b} \notin \boldsymbol{x}_0 + \mathcal{K}_{k-1} \Rightarrow \boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0 \notin A\mathcal{K}_{k-1} \,. \tag{3.4}$$

This means that we can find $\boldsymbol{p}_k \in (A\mathcal{K}_{k-1})^\perp$ such that $\boldsymbol{p}_k^t \boldsymbol{r}_0 \neq 0$; and $\boldsymbol{p}_k^t \boldsymbol{r}_k \neq 0$ follows from the fact that $\boldsymbol{p}_k \in (A\mathcal{K}_{k-1})^\perp$ and a similar argument using (3.3).

From the previous proof, we know that if $\boldsymbol{r}_k \neq 0$ then $\boldsymbol{r}_k \notin A\mathcal{K}_{k-1}$[1]. Following this observation, the *conjugate gradient method* finds $\boldsymbol{p}_k$ by removing from $\boldsymbol{r}_k$ the latter's projection onto $(A\mathcal{K}_{k-1})^\perp$. That is, we write:

$$\boldsymbol{r}_k = \gamma_k \boldsymbol{p}_k + \boldsymbol{z}_k\,, \quad \boldsymbol{z}_k \in A\mathcal{K}_{k-1}\,,\ \boldsymbol{p}_k \in (A\mathcal{K}_{k-1})^\perp \text{ and } \gamma_k \neq 0. \tag{3.5}$$

More details about this decomposition as well as an explicit formula to compute $\gamma_k \boldsymbol{p}_k$ and $\boldsymbol{z}_k$ from a basis of $A\mathcal{K}_{k-1}$ is described in Exercise 1.

Because $\boldsymbol{r}_k - \boldsymbol{r}_0 = A(\boldsymbol{x}_0 - \boldsymbol{x}_k) \in A\mathcal{K}_{k-1}$, it is not difficult to see that the orthogonal decomposition of $\boldsymbol{r}_0$ w.r.t. $A\mathcal{K}_{k-1}$ is given by:

$$\boldsymbol{r}_0 = \gamma_k \boldsymbol{p}_k + (\boldsymbol{z}_k + \boldsymbol{r}_0 - \boldsymbol{r}_k)\,.$$

Now let $\boldsymbol{p}_0 = \boldsymbol{r}_0$, we can then use induction to show that:

$$\mathcal{K}_k = \operatorname{span}(\boldsymbol{r}_0\,,\ A\boldsymbol{r}_0\,,\ \cdots,\ A^k \boldsymbol{r}_0)\,. \tag{3.6}$$

Indeed, the choice of $\boldsymbol{p}_0$ indicates $\mathcal{K}_0 = \operatorname{span}(\boldsymbol{r}_0)$ and now we assume (3.6) is true for $k-1$ ($k \geq 1$), then we have:

$$\boldsymbol{p}_k = \frac{1}{\gamma_k}[\boldsymbol{r}_0 - (\boldsymbol{z}_k + \boldsymbol{r}_0 - \boldsymbol{r}_k)] \in \operatorname{span}(\boldsymbol{r}_0\,,\ A\mathcal{K}_{k-1}) = \operatorname{span}(\boldsymbol{r}_0\,,\ A\boldsymbol{r}_0\,,\ \cdots,\ A^k \boldsymbol{r}_0)\,,$$

and (3.6) for the case k follows immediately.

Next let's take a closer look at $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k = \boldsymbol{b} - A(\boldsymbol{x}_0 + P_{k-1}\boldsymbol{y}_{k-1})$, where $\boldsymbol{y}_{k-1}$ minimize $\phi(\boldsymbol{x}_0 + P_{k-1}\boldsymbol{y})$ over all $\boldsymbol{y} \in \mathbb{R}^k$. The solution to this minimization problem is not difficult to compute:

$$\boldsymbol{y}_{k-1} = (P_{k-1}^t A P_{k-1})^{-1} P_{k-1}^t A \boldsymbol{r}_0\,.$$

It is then computed that:

$$P_{k-1}^t \boldsymbol{r}_k = P_{k-1}^t (\boldsymbol{r}_0 - A P_{k-1} \boldsymbol{y}_{k-1}) = 0\,,$$

or equivalently $\boldsymbol{r}_k \in \mathcal{K}_{k-1}^\perp$. Combining this result and (3.5), we obtain again by induction that $\boldsymbol{r}_0\,, \boldsymbol{r}_1\,, \quad, \boldsymbol{r}_k$ are orthogonal to each other (hence are also linearly independent) and:

$$\mathcal{K}_k = \operatorname{span}(\boldsymbol{r}_0\,,\ \boldsymbol{r}_1\,,\ \cdots,\ \boldsymbol{r}_k)\,. \tag{3.7}$$

There remains an important step of deriving a formula to compute $\boldsymbol{z}_k$ efficiently. Let us assume for now $k \geq 2$. Note that $\boldsymbol{z}_k \in A\mathcal{K}_{k-1}$, one of whose basis is given by AP_{k-1}, hence:

$$\boldsymbol{z}_k = A P_{k-1} \boldsymbol{y}_{k-1}\,, \quad \boldsymbol{y}_{k-1} = \arg\min_{\boldsymbol{y} \in \mathbb{R}^k} ||\boldsymbol{r}_k - A P_{k-1} \boldsymbol{y}||\,.$$

Let us write $\boldsymbol{y} \in \mathbb{R}^k$ and P_{k-1} as

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{w} \\ \beta \end{bmatrix}, \quad \boldsymbol{w} \in \mathbb{R}^{k-1}\,,\ \beta \in \mathbb{R}\,; \qquad P_{k-1} = \begin{bmatrix} P_{k-2} & \boldsymbol{p}_{k-1} \end{bmatrix},$$

[1] Otherwise if $\boldsymbol{r}_k \in A\mathcal{K}_{k-1}$, then for all $\boldsymbol{p} \in (A\mathcal{K}_{k-1})^\perp$ we must have $\boldsymbol{p}^t \boldsymbol{r}_k = 0$, contradiction.

then the vector whose 2-norm is to be minimized is the same as:

$$\boldsymbol{r}_k - AP_{k-1}\boldsymbol{y} = \boldsymbol{r}_k - AP_{k-2}\boldsymbol{w} - \beta A\boldsymbol{p}_{k-1}.$$

Noticing that:

$$\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k = \boldsymbol{b} - A(\boldsymbol{x}_{k-1} + \alpha_{k-1}\boldsymbol{p}_{k-1}) = \boldsymbol{r}_{k-1} - \alpha_{k-1}A\boldsymbol{p}_{k-1}\,,$$

we further compute:

$$\boldsymbol{r}_k - AP_{k-1}\boldsymbol{y} = \boldsymbol{r}_k - AP_{k-2}\boldsymbol{w} - \frac{\beta}{\alpha_{k-1}}(\boldsymbol{r}_{k-1} - \boldsymbol{r}_k) = \left(1 + \frac{\beta}{\alpha_{k-1}}\right)\boldsymbol{r}_k - AP_{k-2}\boldsymbol{w} - \frac{\beta}{\alpha_{k-1}}\boldsymbol{r}_{k-1}\,.$$

Because both $AP_{k-2}\boldsymbol{w}$ and $\boldsymbol{r}_{k-1}$ belongs to $\mathcal{K}_{k-1}$, using the previous result that $\boldsymbol{r}_k \in \mathcal{K}_{k-1}^{\perp}$ we conclude that:

$$||\boldsymbol{r}_k - AP_{k-1}\boldsymbol{y}||^2 = \left(1 + \frac{\beta}{\alpha_{k-1}}\right)^2 ||\boldsymbol{r}_k||^2 + \left|\left|AP_{k-2}\boldsymbol{w} + \frac{\beta}{\alpha_{k-1}}\boldsymbol{r}_{k-1}\right|\right|^2.$$

Let $\boldsymbol{y}_{k-1}$ have the components $\boldsymbol{w}_{k-1}$ and β_{k-1}, then $\boldsymbol{w}_{k-1}$ solves the minimization problem:

$$-\frac{\alpha_{k-1}}{\beta_{k-1}}\boldsymbol{w}_{k-1} = \arg\min_{\boldsymbol{w}\in\mathbb{R}^{k-1}} ||\boldsymbol{r}_{k-1} - AP_{k-2}\boldsymbol{w}||\,,$$

or equivalently:

$$-\frac{\alpha_{k-1}}{\beta_{k-1}}AP_{k-2}\boldsymbol{w}_{k-1} = \arg\min_{\boldsymbol{z}\in A\mathcal{K}_{k-2}} ||\boldsymbol{r}_{k-1} - \boldsymbol{z}|| = \boldsymbol{z}_{k-1} = \boldsymbol{r}_{k-1} - \gamma_{k-1}\boldsymbol{p}_{k-1}\,.$$

This leads to a formula of $\boldsymbol{z}_k$:

$$\boldsymbol{z}_k = AP_{k-2}\boldsymbol{w}_{k-1} + \beta_{k-1}A\boldsymbol{p}_{k-1} = -\frac{\beta_{k-1}}{\alpha_{k-1}}(\boldsymbol{r}_{k-1} - \gamma_{k-1}\boldsymbol{p}_{k-1}) + \beta_{k-1}A\boldsymbol{p}_{k-1} = -\frac{\beta_{k-1}}{\alpha_{k-1}}(\boldsymbol{r}_k - \gamma_{k-1}\boldsymbol{p}_{k-1})\,.$$

In the view of (3.5), $\boldsymbol{p}_k$ is a linear combination of $\boldsymbol{r}_k$ and $\boldsymbol{p}_{k-1}$. This statement is clearly also true for $k < 2$.

Remark 3. *Strictly speaking, we need to show first that $\beta_{k-1} \neq 0$ in the preceding argument. But this is actually fairly straightforward by an argument of contradiction.*

At last, without loss of generality, we can always scale the search direction appropriately so that:

$$\boldsymbol{p}_k = \boldsymbol{r}_k + s_k\boldsymbol{p}_{k-1}\,, \tag{3.8}$$

where the scalar s_k can be evaluated by pre-multiplying both sides of (3.8) by $\boldsymbol{p}_{k-1}^t A$:

$$0 = \boldsymbol{p}_{k-1}^t A\boldsymbol{p}_k = \boldsymbol{p}_{k-1}^t A\boldsymbol{r}_k + s_k\boldsymbol{p}_{k-1}^t A\boldsymbol{p}_{k-1}\,,$$

here the first identity is due to the fact that $\boldsymbol{p}_k$ is A-conjugate to $\mathcal{K}_{k-1}$. Thus:

$$s_k = -\frac{\boldsymbol{p}_{k-1}^t A\boldsymbol{r}_k}{\boldsymbol{p}_{k-1}^t A\boldsymbol{p}_{k-1}}\,. \tag{3.9}$$

In the end, the detailed algorithm for the conjugate gradient method is given by the following:

$$
\begin{aligned}
&\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0\,, \quad k = 0\,; \\
&\text{while } \boldsymbol{r}_k \neq 0\,: \\
&\qquad \text{if } k = 0 \\
&\qquad\qquad \boldsymbol{p}_0 = \boldsymbol{r}_0\,; \\
&\qquad \text{else} \\
&\qquad\qquad s_k = -\boldsymbol{p}_{k-1}^t A\boldsymbol{r}_k / \boldsymbol{p}_{k-1}^t A\boldsymbol{p}_{k-1}\,; \\
&\qquad\qquad \boldsymbol{p}_k = \boldsymbol{r}_k + s_k \boldsymbol{p}_{k-1}\,; \\
&\qquad \text{end} \\
&\qquad \alpha_k = \boldsymbol{p}_k^t \boldsymbol{r}_k / \boldsymbol{p}_k^t A\boldsymbol{p}_k\,; \\
&\qquad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k\,; \\
&\qquad \boldsymbol{r}_{k+1} = \boldsymbol{b} - A\boldsymbol{x}_{k+1}\,; \\
&\qquad k = k+1\,; \\
&\text{end} \\
&\boldsymbol{x} = \boldsymbol{x}_k\,.
\end{aligned}
\tag{3.10}
$$

4 Further Analysis

There are several steps of the algorithm (3.10) that can be improved for computational efficiency. In particular, there are four matrix-vector multiplications in each iteration, one of which can be re-used between adjacent loops; hence algorithm (3.10) has in average three matrix-vector multiplications per iteration. This is the major computational cost with the method.

Now let us derive an equivalent method that only requires one matrix-vector multiplication per iteration. Particularly, pre-multiplying $\boldsymbol{r}_k = \boldsymbol{r}_{k-1} - \alpha_{k-1} A\boldsymbol{p}_{k-1}$ by $\boldsymbol{r}_k^t$ we obtain:

$$
\boldsymbol{r}_k^t \boldsymbol{r}_k = -\alpha_{k-1} \boldsymbol{r}_k^t A\boldsymbol{p}_{k-1}\,;
$$

and pre-multiplying by $\boldsymbol{p}_{k-1}^t = \boldsymbol{r}_{k-1}^t - s_{k-1}\boldsymbol{p}_{k-2}^t$ we obtain:

$$
0 = \boldsymbol{p}_{k-1}^t \boldsymbol{r}_k = (\boldsymbol{r}_{k-1}^t - s_{k-1}\boldsymbol{p}_{k-2}^t)\boldsymbol{r}_{k-1} - \alpha_{k-1}\boldsymbol{p}_{k-1}^t A\boldsymbol{p}_{k-1} = \boldsymbol{r}_{k-1}^t \boldsymbol{r}_{k-1} - \alpha_{k-1}\boldsymbol{p}_{k-1}^t A\boldsymbol{p}_{k-1}\,.
$$

Thus we can compute s_k and α_k instead as:

$$
s_k = -\frac{\boldsymbol{p}_{k-1}^t A\boldsymbol{r}_k}{\boldsymbol{p}_{k-1}^t A\boldsymbol{p}_{k-1}} = \frac{\boldsymbol{r}_k^t \boldsymbol{r}_k}{\boldsymbol{r}_{k-1}^t \boldsymbol{r}_{k-1}}\,, \quad \alpha_k = \frac{\boldsymbol{r}_k^t \boldsymbol{r}_k}{\boldsymbol{p}_k^t A\boldsymbol{p}_k}\,. \tag{4.1}
$$

Finally, the next residual is updated from the previous one by:

$$
\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k A\boldsymbol{p}_k\,, \tag{4.2}
$$

where the product of $A\boldsymbol{p}_k$ is already available in computing α_k. To this end, the following version

of the method only requires one matrix-vector multiplication per iteration:

$$
\begin{aligned}
&\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0\,, \quad k=0\,; \\
&\text{while } \boldsymbol{r}_k \neq 0 : \\
&\qquad \text{if } k=0 \\
&\qquad\qquad \boldsymbol{p}_0 = \boldsymbol{r}_0\,; \\
&\qquad \text{else} \\
&\qquad\qquad s_k = \boldsymbol{r}_k^t \boldsymbol{r}_k / \boldsymbol{r}_{k-1}^t \boldsymbol{r}_{k-1}\,; \\
&\qquad\qquad \boldsymbol{p}_k = \boldsymbol{r}_k + s_k \boldsymbol{p}_{k-1}\,; \\
&\qquad \text{end} \\
&\qquad \alpha_k = \boldsymbol{r}_k^t \boldsymbol{r}_k / \boldsymbol{p}_k^t A \boldsymbol{p}_k\,; \\
&\qquad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k\,; \\
&\qquad \boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k A \boldsymbol{p}_k\,; \\
&\qquad k = k+1\,; \\
&\text{end} \\
&\boldsymbol{x} = \boldsymbol{x}_k\,.
\end{aligned}
\tag{4.3}
$$

Using the structure of $\mathcal{K}_k$, we see that with exact arithmetics the algorithm (4.3) converges in at most n steps; and if $A = I + B$ such that $\text{rank} B = r$, it converges in at most $r+1$ steps.

Finally, we state without proof an estimate on the error of the conjugate gradient method:

$$
||\boldsymbol{x} - \boldsymbol{x}_k||_A \leq 2||\boldsymbol{x} - \boldsymbol{x}_0||_A \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \right)^k , \tag{4.4}
$$

where $\kappa = \kappa_2(A)$ is the condition number of A in the induced L^2-norm. More details can be found in David G. Luenberger's work [1].

Exercises

Exercise 1. *Let $\mathcal{K}$ be a k-dimensional subspace of $\mathbb{R}^n$ with a basis $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$. Then the orthogonal complement of $\mathcal{K}$ is defined as:*

$$
\mathcal{K}^{\perp} = \{ \boldsymbol{w} \in \mathbb{R}^n \,:\, \boldsymbol{w}^t \boldsymbol{v} = 0 \;\; \forall \boldsymbol{v} \in \mathcal{K} \}\,. \tag{4.5}
$$

*Let $\boldsymbol{u} \in \mathbb{R}^n$ be arbitrary, its **orthogonal decomposition** with respect to $\mathcal{K}$ is:*

$$
\boldsymbol{u} = \boldsymbol{w} + \boldsymbol{v}\,, \quad \text{such that } \boldsymbol{w} \in \mathcal{K}^{\perp} \text{ and } \boldsymbol{v} \in \mathcal{K}\,. \tag{4.6}
$$

(i) Show that this decomposition is unique.

Usually we denote this unique $\boldsymbol{v}$ by $\mathcal{P}\boldsymbol{u}$ and $\boldsymbol{w}$ by $(\mathcal{I} - \mathcal{P})\boldsymbol{u}$ where $\mathcal{P} : \mathbb{R}^n \to \mathcal{K}$ is known as the projection of $\mathbb{R}^n$ onto $\mathcal{K}$ and $\mathcal{I} : \mathbb{R}^n \to \mathbb{R}^n$ is the identity map.

(ii) Prove that $\boldsymbol{v} = \mathcal{P}\boldsymbol{u}$ minimizes the distance between any vector of $\mathcal{K}$ and $\boldsymbol{u}$, i.e.,

$$
\boldsymbol{v} = \mathcal{P}\boldsymbol{u} = \arg\min_{\boldsymbol{v}' \in \mathcal{K}} \big|\big| \boldsymbol{v}' - \boldsymbol{u} \big|\big|\,. \tag{4.7}
$$

***Hint**: We just need to show that $||\boldsymbol{v}'-\boldsymbol{u}|| \geq ||\boldsymbol{v}-\boldsymbol{u}||$ for all $\boldsymbol{v}' \in \mathcal{K}$. This is rewritten as $||(\boldsymbol{v}'-\boldsymbol{v})-\boldsymbol{w}|| \geq ||\boldsymbol{w}||$. To proceed, use $\boldsymbol{w} \in \mathcal{K}^{\perp}$ to show that $||(\boldsymbol{v}'-\boldsymbol{v})-\boldsymbol{w}||^2 = ||\boldsymbol{v}'-\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2$.*

Now let $V = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_k] \in \mathbb{R}^{n\times k}$, then any vector of $\mathcal{K}$ can be written as $V\boldsymbol{y}$ with some $\boldsymbol{y} \in \mathbb{R}^k$. (iii) Use this representation and (4.7) to show:

$$\mathcal{P}\boldsymbol{u} = V(V^tV)^{-1}V^t\boldsymbol{u}\,. \tag{4.8}$$

References

[1] David G. Luenberger. Introduction to Linear and Nonlinear Programming. Addison Wesley Publishing Company, 1973.