Numerical Optimization

Exercise III

Jonathan Balzer

Geometric Modeling and Scientific Visualization Center
King Abdullah University of Science and Technology

CG method in a nutshell I

Cayley-Hamilton tells us:

$$p_{\mathrm{c}}^{n}(\mathbf{A}) = \mathbf{A}^0 + \alpha_0 \mathbf{A} + \ldots \alpha_{n-1}\mathbf{A}^n = \mathbf{0} \in \mathbb{R}^{n\times n}$$
$$\Leftrightarrow \quad \mathbf{A}^{-1} = -\alpha_0 \mathbf{I} - \alpha_2 \mathbf{A} - \ldots - \alpha_{n-1}\mathbf{A}^{n-1} =: p^{n-1}(\mathbf{A})$$

The "path" from $\boldsymbol{x}_0$ to the solution $\boldsymbol{x}^*$ can be decomposed into a finite number of "segments"

$$\boldsymbol{x}^* - \boldsymbol{x}_0 = \mathbf{A}^{-1}(\boldsymbol{b} - \mathbf{A}\boldsymbol{x}_0) = p^{n-1}(\mathbf{A})\boldsymbol{r}_0$$

or more generally

$$\boldsymbol{x}^* - \boldsymbol{x}_0 \in \mathrm{span}\{\boldsymbol{r}_0, \mathbf{A}\boldsymbol{r}_0, \ldots, \mathbf{A}^{n-1}\boldsymbol{r}_0\} = \mathcal{K}_n(\mathbf{A}, \boldsymbol{r}_0)$$

(*Krylov space* of dimension n)

Solution of $\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow$ finding a favorable basis in $\mathcal{K}_n(\mathbf{A}, \boldsymbol{r}_0)$, step sizes α_k = coefficients in this basis

CG method in a nutshell II

Idea:

Build basis iteratively by $\mathbf{A}$-orthogonalization of residuals $\boldsymbol{r}_k$!

Updating the descent direction:

1. actual $\mathbf{A}$-orthogonal basis $B_{k-1} = \{\boldsymbol{p}_0, \dots, \boldsymbol{p}_{k-1}\}$
2. $\boldsymbol{p}_{\boldsymbol{k}} = -\boldsymbol{r}_k$ and hence $B_k = \{\boldsymbol{p}_0, \dots, \boldsymbol{p}_{k-1}, -\boldsymbol{r}_k\}$
3. Gram-Schmidt:

$$\boldsymbol{p}_k = -\boldsymbol{r}_k + \sum_{i=0}^{k-1} \underbrace{\boldsymbol{r}_k^\top \mathbf{A} \boldsymbol{p}_i}_{\beta_{i+1}} \boldsymbol{p}_i$$

Trick: Pre-multiply by $\boldsymbol{p}_{k-1}^\top \mathbf{A}$ so that

$$0 =: -\boldsymbol{p}_{k-1}^\top \mathbf{A} \boldsymbol{r}_k + \beta_k \boldsymbol{p}_{k-1}^\top \mathbf{A} \boldsymbol{p}_{k-1}$$

from which β_k is obtained easily.

Exercises

1. Given a vector $\boldsymbol{v} = (0, 1, 0)^\top$. Calculate a vector $\boldsymbol{v}^\perp$ which is $\mathbf{A}$-orthogonal to $\boldsymbol{v}$ with respect to

$$\mathbf{A} = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}.$$

Exercises

2. Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

(i) Find a basis in which $\mathbf{A}$ has diagonal form to show that it fails to be positive definite.

(ii) Construct a matrix $\mathbf{E}$ such that $\mathbf{A} + \mathbf{E}$ has only strictly positive eigenvalues.

Exercises

3. Given the function, $f : \mathbb{R}^3 \mapsto \mathbb{R}$,

$$f(x_1, x_2, x_3) = \exp(x_1 x_2) + \frac{x_2^2}{x_3 - 1}$$

a point $\boldsymbol{x} = (0, 2, 3)^\top \in \mathbb{R}^3$ and a direction $\boldsymbol{p} = (-1, 1, -1)^\top \in \mathbb{R}^3$.

(i) Construct the computational graph of f.

(ii) Evaluate $f(\boldsymbol{x})$ recording all intermediate values in the graph.

(iii) Calculate $\nabla f(\boldsymbol{x})$ by AD in reverse mode and $\nabla f(\boldsymbol{x})^\top \boldsymbol{p}$ in forward mode.

Convex functions revisited

- $f : \mathbb{R}^n \supseteq \Omega \mapsto \mathbb{R}$ convex if Ω is convex and

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

  for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ and $\lambda \in [0, 1]$

- f strictly convex if above inequality is strict

- f concave $\Leftrightarrow -f$ convex

Conditions for convexity

- first-order condition ($f \in C^1$ and Ω convex):

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x})$$

- second-order condition ($f \in C^2$ and Ω convex):

$$\nabla^2 f \geq 0$$

Convexity preserving operations

1. restriction to affine spaces: $f : \mathbb{R}^n \mapsto \mathbb{R}$ convex iff

   $$g(t) := f(\boldsymbol{x}_0 + t\boldsymbol{v})$$

   convex in t for any $\boldsymbol{x}_0 \in \Omega, \boldsymbol{v} \in \mathbb{R}^n$

2. affine transformations: if $f(\boldsymbol{x})$ is convex so is $f(\mathbf{A}\boldsymbol{x} + \boldsymbol{t})$ for any $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{t} \in \mathbb{R}^n$

3. scaling: αf convex if f convex and $\alpha \in \mathbb{R}_{\geq 0}$

4. sum: f_1, f_2 convex $\Rightarrow$ $f_1 + f_2$ convex

5. pointwise maximum: $f_1, \ldots, f_m$ convex $\Rightarrow$ $f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})\}$ convex

6. composition: g convex, h convex and non-decreasing $\Rightarrow$ $f = h \circ g(\boldsymbol{x})$ convex

Exercises

4.

(i) Show that $f(x) = \exp(-x)$ is convex but $f(x) = \exp(-x^2)$ is not.

(ii) Proof statements 3. and 4. from above.