Prof. Dr. Dr. h.c. Florian Rupp
Akaki Matcharashvili

General Descent Methods – CE Solutions

Short Recap of the Lecture Topics:

General Descent Methods & Admissible Descent Directions:

- Let $\{s^k\}_{k\in\mathbb{N}_0}$ be the sequence of descent directions generated by the general descent method. A sub-sequence $\{s^\ell\}_{\ell\in L\subset\mathbb{N}_0}$ is called **admissible**, if the condition C1 holds

$$\nabla f(x^k)^T s^k \ < \ 0 \qquad \text{for all } k\in\mathbb{N}_0$$

(i.e. s^k is a descent direction) together with the condition C2:

$$\left\{\frac{\nabla f(x^\ell)^T s^\ell}{\|s^\ell\|}\right\}_{\ell\in L} \overset{\ell\to\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad \left\{\nabla f(x^\ell)\right\}_{\ell\in L} \overset{\ell\to\infty}{\longrightarrow} 0$$

- **Angle condition (AC)**: For an $0<\eta<1$ and for all $\ell\in L$ it holds that

$$\cos\Big(\angle(-\nabla f(x^\ell), s^\ell)\Big) \ = \ \frac{-\nabla f(x^\ell)^T s^\ell}{\|\nabla f(x^\ell)\|\,\|s^\ell\|} \ \geq \ \eta\,.$$

- **Generalized angle condition (GAC)**: For a suitable chosen function $\phi:[0,\infty)\to[0,\infty)$ that is continuous in 0 with $\phi(0)=0$ it holds for all $\ell\in L$ that

$$\|\nabla f(x^\ell)\| \ \leq \ \phi\left(\frac{\nabla f(x^\ell)^T s^\ell}{\|s^\ell\|}\right)\,.$$

- **Theorem:** (AC $\Rightarrow$ GAC $\Rightarrow$ C2) Let $f:\mathbb{R}^n\to\mathbb{R}$ be continuously differentiable and $\{s^\ell\}_{\ell\in L}$ be a sub-sequence of descent directions (i.e. $\nabla f(x^\ell)^T s^\ell<0$ for all $\ell\in L$) generated by the general descent method.

  1. If $\{s^\ell\}_{\ell\in L}$ fulfills the angle condition AC then it also fulfills the generalized angle condition GAC.
  2. If $\{s^\ell\}_{\ell\in L}$ fulfills the general angle condition GAC then it is also admissible (and fulfills C2).

  Here, η and ϕ must (of course) be independent of $\ell\in L$.

Admissible Step Size Rules:

- Let $\{\sigma_k\}_{k\in\mathbb{N}_0}$ be the sequence of step sizes generated by the general descent method. A sub-sequence $\{\sigma_\ell\}_{\ell\in L\subset\mathbb{N}_0}$ is called **admissible**, if

$$f(x^\ell+\sigma_\ell s^\ell) \ \leq \ f(x^\ell) \qquad \text{for all } \ell\in L$$

and

$$f(x^\ell+\sigma_\ell s^\ell)-f(x^\ell) \overset{\ell\to\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad \left\{\frac{\nabla f(x^\ell)^T s^\ell}{\|s^\ell\|}\right\}_{\ell\in L} \overset{\ell\to\infty}{\longrightarrow} 0$$

- Let s^k be a descent direction of f at x^k. The step size σ_k is called **efficient**, if there is a constant $\theta>0$ such that

$$f(x^k+\sigma_k s^k) \ \leq \ f(x^k)-\theta\left(\frac{\nabla f(x^k)^T s^k}{\|s^k\|}\right)^2\,.$$

- **Lemma:** Let $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable. Let the sequences $\{x^k\}_{k\in\mathbb{N}_0}$, $\{s^k\}_{k\in\mathbb{N}_0}$, and $\{\sigma_k\}_{k\in\mathbb{N}_0}$ be generated by the general descent method with $f(x^k + \sigma_k s^k) \le f(x^k)$ for all $k \in \mathbb{N}_0$. Let $\{\sigma_\ell\}_{\ell\in L\subset\mathbb{N}_0}$ be a sub-sequence for wich all σ_ℓ, $\ell \in L$, are efficient, then the step size sub-sequence $\{\sigma_\ell\}_{\ell\in L\subset\mathbb{N}_0}$ is admissible.

- **Admissibility of the Armijo Step Sizes:** Let $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable. Let the sequences $\{x^k\}_{k\in\mathbb{N}_0}$, $\{s^k\}_{k\in\mathbb{N}_0}$, and $\{\sigma_k\}_{k\in\mathbb{N}_0}$ be generated by the general descent method. Moreover, let there be

  - a bounded sub-sequence $\{x^\ell\}_{\ell\in L\subset\mathbb{N}_0}$ of iterates (e.g. as $\{x^\ell\}_{\ell\in L}$ converges), and
  - a strictly monotonously increasing function $\phi : [0,\infty) \to [0,\infty)$ such that the sub-sequence of descent directions $\{s^\ell\}_{\ell\in L}$ corresponding to $\{x^\ell\}_{\ell\in L}$ satisfies

  $$\|s^\ell\| \;\ge\; \phi\left(-\frac{\nabla f(x^\ell)^T s^\ell}{\|s^\ell\|}\right) \qquad \text{for all } \ell \in L\,.$$

  Then, the corresponding sub-sequence of step sizes $\{\sigma_\ell\}_{\ell\in L}$ generated by the Armijo step size rule is admissible.

- **Global Convergence Theorem for a Descent Method:** Let $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable. Let $x^0 \in \mathbb{R}^n$ arbitrary and the sequences $\{x^k\}_{k\in\mathbb{N}_0}$, $\{s^k\}_{k\in\mathbb{N}_0}$, and $\{\sigma_k\}_{k\in\mathbb{N}_0}$ be generated by the general descent method. Moreover, let $\overline{x} \in \mathbb{R}^n$ be an accumulation point of $\{x^k\}_{k\in\mathbb{N}_0}$ with a sub-sequence $\{x^\ell\}_{\ell\in L\subset\mathbb{N}_0} \to \overline{x}$ for $\ell \to \infty$, $\ell \in L$.

  If $\{s^\ell\}_{\ell\in L}$, and $\{\sigma_\ell\}_{\ell\in L}$ are admissible, then $\overline{x}$ is a stationary point, i.e. $\nabla f(\overline{x}) = 0$.

<u>**Solved Central Exercise Problems:**</u>

Exercise 5.1: Criteria for Admissibility — Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable, and (x^k), (s^k) and (σ_k) be generated by the general descent method. Prove the following assertions:

a) A sub-sequence of search directions $\{s^{k_j}\}$ is admissible, if there exists a strictly monotonically increasing function $\varphi : [0,\infty) \to [0,\infty)$ with $\varphi(0) = 0$, that satisfies

$$-\nabla f(x^{k_j})^T s^{k_j} \;\ge\; \varphi\Big(\|\nabla f(x^{k_j})\|\Big)\|s^{k_j}\| \quad \forall j\,.$$

b) A sub-sequence of step sizes (σ_{k_j}) is admissible, if for any $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$, such that the following implication holds:

$$\begin{aligned} &\frac{-\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|} \;\ge\; \varepsilon \text{ for infinitely many } j \\ &\Longrightarrow\; f(x^{k_j}) - f(x^{k_j} + \sigma_{k_j} s^{k_j}) \;\ge\; \delta(\varepsilon) \text{ for infinitely many } j. \end{aligned}$$

Solution:

ad a) The first condition of admissibility, is that s^k is a descent direction, which is always true for a descent method, i.e. it holds:

$$\nabla f(x^k)^T s^k \;<\; 0 \quad \forall k$$

We will show only the second condition, i.e.:

$$\left\{\frac{\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|}\right\}_j \to 0 \quad\Longrightarrow\quad \Big\{\nabla f(x^{k_j})\Big\}_j \to 0\,.$$

We show this by contraposition. Thus, let $\big\{\nabla f(x^{k_j})\big\}_j \not\to 0$. Then there is $\varepsilon > 0$ and a sub-sequence of $\{k_j\}$ (for simplicity denoted by k_j again) such that

$$\|\nabla f(x^{k_j})\| \;\ge\; \varepsilon \quad \forall j\,.$$

By the strictly monotonous growth of φ and $\varphi(0) = 0$:

$$\frac{-\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|} \;\geq\; \varphi(\|\nabla f(x^{k_j})\|) \;\geq\; \varphi(\varepsilon) \;>\; \varphi(0) \;=\; 0 \quad \forall\, j\,.$$

This shows $\left\{\frac{-\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|}\right\}_j \not\to 0$, concluding the contraposition.

ad b) Since the sequence $\{\sigma_k\}$ was generated by a descent method, it holds

$$f(x^k + \sigma_k s^k) \;<\; f(x^k) \quad \forall\, k\,.$$

Thus the first condition for admissibility is satisfied.
We show the second condition, i.e.

$$f(x^{k_j} + \sigma_{k_j} s^{k_j}) - f(x^{k_j}) \;\to\; 0 \quad \Longrightarrow \quad \left\{\frac{\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|}\right\}_j \;\to\; 0\,,$$

again by contraposition. Let $\left\{\frac{\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|}\right\}_j \not\to 0$. Since s^{k_j} are descent directions, there is an $\varepsilon > 0$ and a sub-sequence of $\{k_j\}$ (for simplicity again denoted by k_j), with

$$\frac{-\nabla f(x^{k_j})^T s^{k_j}}{\|s^{k_j}\|} \;\geq\; \varepsilon \quad \forall\, j \quad \text{(thus for infinitely many } j).$$

Thus by assumption, there is a $\delta > 0$ with

$$f(x^{k_j}) - f(x^{k_j} + \sigma_{k_j} s^{k_j}) \;\geq\; \delta \quad \text{for infinitely many } j,$$

and thus it holds $f(x^{k_j} + \sigma_{k_j} s^{k_j}) - f(x^{k_j}) \not\to 0$, which concludes the contraposition.

Exercise 5.2: Curry Step Size Rule — Let $f : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable, and $x^0 \in \mathbb{R}^n$ such that the level-set $\mathcal{N}_f(x^0) = \{x \in \mathbb{R}^n \,:\, f(x) \leq f(x^0)\}$ is compact. Let further $x \in \mathcal{N}_f(x^0)$ and $s \in \mathbb{R}^n$ a descent direction of f in x. The **Curry Step Size Rule** computes $\overline{\sigma} > 0$ as the smallest positive stationary point of the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by $\phi(\sigma) := f(x + \sigma s)$:

$$\overline{\sigma} \;:=\; \min\left\{\sigma > 0 \,:\, \phi'(\sigma) = 0\right\}\,.$$

a) Show well-posedness of the Curry step size rule.
Hint: Assume first, that there is no positive stationary point of ϕ, and use the mean value theorem to construct a contradiction to the compactness of $\mathcal{N}_f(x^0)$. Further show, that under all stationary points of ϕ there is actually a smallest one.

b) Show, using the Intermediate Value Theorem, that there is a smallest $0 < \hat{\sigma} < \overline{\sigma}$ such that $\nabla f(x + \hat{\sigma} s)^T s = \frac{1}{2} \nabla f(x)^T s$.

c) Show using part b): $f(x) - f(x + \overline{\sigma} s) \geq -\frac{1}{2}\hat{\sigma} \nabla f(x)^T s$.

d) Use compactness of $\mathcal{N}_f(x^0)$ and continuity of ∇f to show with help of part b), that for any $\varepsilon > 0$ there is a $\delta(\varepsilon) > 0$, independent of x and s, such that

$$-\frac{1}{2}\frac{\nabla f(x)^T s}{\|s\|} \;\geq\; \varepsilon \quad \Longrightarrow \quad \|\hat{\sigma} s\| \;\geq\; \delta(\varepsilon)\,.$$

e) Use parts c) and d), to show that any sub-sequence of step sizes $\{\sigma_k\}_K$, generated by the Gradient Descent Method with starting point x^0 and the Curry step size rule is admissible.

Solution:

ad a) Assume, that there is no positive stationary point of ϕ. By $\phi'(0) = \nabla f(x)^T s < 0$ and by continuity for all $\sigma > 0$, it holds that $\phi'(\sigma) = \nabla f(x + \sigma s)^T s < 0$. Thus also for all $\sigma > 0$:

$$f(x + \sigma s) - f(x) \;=\; \phi(\sigma) - \phi(0) \;=\; \sigma \phi'(\xi) \;=\; \sigma \nabla f(x + \xi s)^T s \;<\; 0 \quad \text{with } \xi \in (0, \sigma)\,,$$

and thus $x + \sigma s \in \mathcal{N}_f(x^0)$. This contradicts the compactness, more specifically the boundedness of the level-set. It remains to show, that there is a smallest positive stationary point of ϕ. The set $M = \{\sigma \geq 0 \,:\, \phi'(\sigma) = 0\}$ is bounded from below by 0, thus possesses an infimum $\overline{\sigma} \geq 0$. By the closedness of M (note ϕ' is continuous!) the infimum $\overline{\sigma}$ is an element of the set. By $\phi'(0) < 0$ it holds $\overline{\sigma} > 0$, and thus the assertion is shown.

ad b) The derivative $\phi'(\sigma) = \nabla f(x + \sigma s)^T s$ is continuous by assumption. Since

$$\phi'(0) \;=\; \nabla f(x)^T s \;<\; \tfrac{1}{2}\nabla f(x)^T s \;<\; 0 \;=\; \phi'(\overline{\sigma})$$

by the Intermediate Value Theorem, there is an $\hat{\sigma} \in (0, \overline{\sigma})$ with $\phi'(\hat{\sigma}) = \frac{1}{2}\nabla f(x)^T s$. As in part a), from the closedness of the set of candidates, there is a smallest $\hat{\sigma} > 0$ with $\phi'(\hat{\sigma}) = \frac{1}{2}\nabla f(x)^T s$.

ad c) For all $\sigma \in [0, \overline{\sigma}]$ it holds $\phi'(\sigma) \leq 0$. Thus ϕ is monotonically decreasing in $[0, \overline{\sigma}]$. From this, we directly conclude:

$$f(x) - f(x + \overline{\sigma} s) \;\geq\; f(x) - f(x + \overline{\sigma} s)\,.$$

Further the Mean Value Theorem gives

$$f(x) - f(x + \hat{\sigma} s) \;=\; -\hat{\sigma}\nabla f(x + \xi s)^T s \;=\; -\hat{\sigma}\phi'(\xi)$$

with $\xi \in [0, \hat{\sigma}]$. Since $\hat{\sigma}$ was the smallest σ with $\phi'(\sigma) = \frac{1}{2}\phi'(0)$ and ϕ' is continuous, it holds $\phi'(\sigma) \leq \phi'(\hat{\sigma})$ for all $\sigma \in [0, \hat{\sigma}]$. Hence, part b) gives

$$-\hat{\sigma}\nabla f(x + \xi s)^T s \;\geq\; -\hat{\sigma}\nabla f(x + \hat{\sigma} s)^T s \;=\; -\tfrac{1}{2}\hat{\sigma}\nabla f(x)^T s\,.$$

ad d) By the compactness of the levelset, it follows even, that ∇f is uniformly continuous in $\mathcal{N}_f(x^0)$. We use, that

$$-\tfrac{1}{2}\nabla f(x)^T s \;=\; \left(\nabla f(x + \hat{\sigma} s) - \nabla f(x)\right)^T s$$

and obtain by Cauchy-Schwarz inequality

$$\epsilon \;\leq\; -\frac{1}{2}\frac{\nabla f(x)^T s}{\|s\|} \;\leq\; \|\nabla f(x + \hat{\sigma} s) - \nabla f(x)\|\,.$$

Since $x \in \mathcal{N}_f(x^0)$ as well as $x + \hat{\sigma} s \in \mathcal{N}_f(x) \subset \mathcal{N}_f(x^0)$ by part c), the uniform continuity of ∇f on $\mathcal{N}_f(x^0)$ implies the existence of some $\delta(\epsilon) > 0$ with $\|x - (x + \hat{\sigma} s)\| = \|\hat{\sigma} s\| \geq \delta(\epsilon)$.

ad e) We show

$$\left\{\frac{\nabla f(x^k)^T s^k}{\|s^k\|}\right\}_K \not\to 0 \quad\Longrightarrow\quad f(x^k + \sigma_k s^k) - f(x^k) \;\not\to\; 0\,.$$

If there holds $\left\{\frac{\nabla f(x^k)^T s^k}{\|s^k\|}\right\}_K \not\to 0$, then there is a sub-sequence $\{x^k\}_{K'}$ of $\{x^k\}_K$ and $\epsilon > 0$ with

$$-\frac{1}{2}\frac{\nabla f(x^k)^T s^k}{\|s^k\|} \;\geq\; \epsilon \quad \forall\, k \,\in\, K'\,.$$

Thus from c) and d) we deduce, that

$$f(x^k) - f(x^k + \sigma_k s^k) \;\geq\; -\tfrac{1}{2}\hat{\sigma}_k\frac{\nabla f(x^k)^T s^k \|s^k\|}{\|s^k\|} \;\geq\; \epsilon\|\hat{\sigma}_k s^k\| \;\geq\; \epsilon\delta(\epsilon) \;>\; 0 \quad \forall\, k \,\in\, K'\,.$$

Hence it holds $f(x^k + \sigma_k s^k) - f(x^k) \not\to 0$ for $k \to \infty$.