Model Based Conditional Gradient Method with Armijo-like Line Search

— Supplementary Material —

A. Proofs

A.1. Proof of Proposition 3.3

For a fixed $k \in \mathbb{N}$, we abbreviate $\gamma = \gamma_k$. Using Assumption 2, we have

$$f(x_{k+1}) - f(x_k) \le f_{x_k}(x_{k+1}) - f_{x_k}(x_k) + \omega(\|x_{k+1} - x_k\|)\,.$$

From $\|x_{k+1} - x_k\| = \gamma \|y_k - x_k\|$ and the definition of a growth function it follows that $\omega(\|x_{k+1} - x_k\|) = o(\gamma)$. The convexity of the model function f_{x_k} gives us

$$f_{x_k}(x_{k+1}) - f_{x_k}(x_k) \le \gamma(f_{x_k}(y_k) - f_{x_k}(x_k))\,.$$

Now, we argue by contradiction. Suppose that for any $\tilde{\gamma} > 0$ there exists $\gamma \in (0, \tilde{\gamma})$ such that (ALS) does not hold, which yields the following calculation

$$\begin{aligned} -\gamma\rho\Delta(x_k, y_k) &< f(x_{k+1}) - f(x_k) \\ &\le \gamma(f_{x_k}(y_k) - f_{x_k}(x_k)) + o(\gamma) \\ &= -\gamma\Delta(x_k, y_k) + o(\gamma)\,. \end{aligned}$$

Dividing the inequality by γ, we obtain

$$0 < (1-\rho)\Delta(x_k, y_k) < o(\gamma)/\gamma\,,$$

which is a contradiction for sufficiently small $\tilde{\gamma}$. □

A.2. Proof of Proposition 3.4

The result is shown by Fermat's rule in the following lemma.

Lemma A.1. *Let $\tilde{x} \in C$. Then,*

$$\widehat{\partial} f(\tilde{x}) = \partial f_{\tilde{x}}(\tilde{x})\,,$$

and

$$0 \in \partial f_{\tilde{x}}(\tilde{x}) \quad \Leftrightarrow \quad \Delta(\tilde{x}, x) \le 0 \;\forall x \in C\,.$$

Proof. Let $v \in \widehat{\partial} f(\tilde{x})$, then

$$f(x) \ge f(\tilde{x}) + \langle v, x - \tilde{x}\rangle + o(\|x - \tilde{x}\|) \quad \forall x \in C$$

and, this implies, by the model assumption for all $x \in C$:

$$f_{\tilde{x}}(x) + \omega(\|x - \tilde{x}\|) \ge f_{\tilde{x}}(\tilde{x}) + \langle v, x - \tilde{x}\rangle + o(\|x - \tilde{x}\|)\,.$$

Since $\omega(t) = o(t)$, we conclude that

$$f_{\tilde{x}}(x) \ge f_{\tilde{x}}(\tilde{x}) + \langle v, x - \tilde{x}\rangle + o(\|x - \tilde{x}\|)\,, \quad \forall x \in C\,.$$

Now, we fix a point $\bar{x} \in C$ and consider $x = \tilde{x} + \tau(\bar{x} - \tilde{x})$ for $\tau \in (0, 1]$. Then, by convexity of C and the model function $f_{\tilde{x}}$, we obtain

$$f_{\tilde{x}}(\tilde{x}) + \tau(f_{\tilde{x}}(\bar{x}) - f_{\tilde{x}}(\tilde{x})) \ge f_{\tilde{x}}(\tilde{x}) + \tau\langle v, \bar{x} - \tilde{x}\rangle + o(\tau\|\bar{x} - \tilde{x}\|)\,.$$

Subtracting $f_{\tilde{x}}(\tilde{x})$, dividing by τ, and considering $\tau \searrow 0$, and, using the fact that this consideration was independent of the choice of $\bar{x}$, we conclude that $v \in \partial f_{\tilde{x}}(\tilde{x})$. The converse direction follows easily.

The second part of the statement is Fermat's rule (Theorem 16.2 in (Bauschke & Combettes, 2011)) for convex functions. □

A.3. Proof of Theorem 3.6

We prove the result in three steps.

Convergence of objective values. The monotonicity and convergence of $(f(x_k))_{k\in\mathbb{N}}$ follows directly from (ALS) and the boundedness of f from below.

Vanishing model improvement. From (ALS) and convergence of $(f(x_k))_{k\in\mathbb{N}}$, we infer that $\gamma_k \Delta(x_k, y_k) \to 0$, since

$$0 \le \rho\gamma_k\Delta(x_k, y_k) \le f(x_k) - f(x_{k+1}) \to 0\,.$$

We deduce boundedness of $(\Delta(x_k, y_k))_{k\in\mathbb{N}}$ by

$$\begin{aligned} 0 &\le \Delta(x_k, y_k) \\ &= f_{x_k}(x_k) - f_{x_k}(y_k) \le f(x_k) - f_{x_k}(\hat{y}_k) \\ &\le f(x_0) - f(\hat{y}_k) + \omega(\|\hat{y}_k - x_k\|) \\ &\le f(x_0) - \inf_{x\in C} f(x) + \omega(\operatorname{diam}(C)) < +\infty\,. \end{aligned}$$

Let Δ^* be an arbitrary limit point of $(\Delta(x_k, y_k))_{k\in\mathbb{N}}$, that is $\Delta(x_k, y_k) \to \Delta^*$ as $k \overset{K}{\to} \infty$ for some $K \subset \mathbb{N}$, where $k \overset{K}{\to} \infty$ abbreviates $k \to \infty$ with $k \in K$.

Suppose $\Delta^* > 0$. Then $\gamma_k \to 0$ as $k \overset{K}{\to} \infty$. For sufficiently large k, the line search procedure in Algorithm 2 reduces

γ_k/δ to γ_k, i.e., (ALS) is violated before multiplying with δ:

$$-\frac{\gamma_k}{\delta}\rho\Delta(x_k, y_k) < f(x_k + \tfrac{\gamma_k}{\delta}(y_k - x_k)) - f(x_k)\,.$$

Analogously to the proof of Proposition 3.3, we conclude

$$\begin{aligned}-\frac{\gamma_k}{\delta}\rho\Delta(x_k, y_k) &< \frac{\gamma_k}{\delta}(f_{x_k}(y_k) - f_{x_k}(x_k)) + o(\gamma_k/\delta)\\ &= -\frac{\gamma_k}{\delta}\Delta(x_k, y_k) + o(\gamma_k/\delta)\,.\end{aligned}$$

Dividing both sides by $\frac{\gamma_k}{\delta}$ results in $(1-\rho)\Delta(x_k, y_k) < o(\gamma_k)/\gamma_k$ and considering $\gamma_k \to 0$ for $k \overset{K}{\to} \infty$ yields a contradiction, since $\rho \in (0, 1)$. Therefore $\Delta(x_k, y_k) \to 0$ for $k \to \infty$.

Convergence to a stationary point. The following relation holds for all $x \in C$:

$$\begin{aligned}\Delta(x_k, y_k) &= \Delta(x_k, \hat{y}_k) + f_{x_k}(\hat{y}) - f_{x_k}(y_k)\\ &\geq f_{x_k}(x_k) - f_{x_k}(x) - \varepsilon_k \\ &\geq f(x_k) - f(x) - \omega(\|x_k - x\|) - \varepsilon_k\,,\end{aligned} \tag{1}$$

where the first inequality follows from Assumption 3 and the second from Assumption 2. Taking the limit $k \overset{K}{\to} \infty$ on both sides, using $\Delta(x_k, y_k) \to 0$ for $k \to \infty$, lower semi-continuity of f and continuity of ω, we arrive at

$$f(x) \geq f(\tilde{x}) - \omega(\|\tilde{x} - x\|)\,, \quad \forall x \in C\,,$$

where $\tilde{x} \in C$ due to compactness of C. As $\tilde{x} \in C$ and $\omega(t) = o(t)$, we deduce that

$$\liminf_{\substack{x \to \tilde{x}\\ x \neq \tilde{x}}} \frac{f(x) - f(\tilde{x}) - \langle 0, x - \tilde{x}\rangle}{\|x - \tilde{x}\|} \geq 0\,.$$

which by definition means that $0 \in \widehat{\partial} f(\tilde{x})$.

Moreover, using $x = \tilde{x}$ in (1), taking the limit $k \overset{K}{\to} \infty$ and using lower semi-continuity of f, we deduce

$$f(\tilde{x}) \geq \limsup_{k \overset{K}{\to} \infty} f(x_k) \geq \liminf_{k \overset{K}{\to} \infty} f(x_k) \geq f(\tilde{x})\,,$$

hence $f(x_k) \to f(\tilde{x})$ as $k \overset{K}{\to} \infty$. By convergence of $(f(x_k))_{k\in\mathbb{N}}$, we also have $f(x_k) \to f(\tilde{x})$ for $k \to \infty$. $\square$

References

Bauschke, H. H. and Combettes, P. L. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.