Direct Search: overview and theory

Optimization of Complex Systems – March 6th 2026

Andrea Brilli

Sapienza University of Rome

Direct Search

Input: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\alpha_{\min} > 0$, $\theta \in (0,1)$, $\gamma > 0$

$k \leftarrow 0$
while $\alpha_k \geq \alpha_{\min}$ **do**
    $k \leftarrow k + 1$
    Generate a set of poll directions $\mathcal{D}_k$
    if $d \in \mathcal{D}_k$ exists such that $f(x_k + \alpha_k d) \leq f(x_k) - \gamma \alpha_k^2$ **then**
        $x_{k+1} \leftarrow x_k + \alpha_k d$, $\alpha_{k+1} \leftarrow \alpha_k$
    else
        $x_{k+1} \leftarrow x_k$, $\alpha_{k+1} \leftarrow \theta \alpha_k$
    end if
end while
return x_k

Key differences from coordinate search:

- Directions d_k can vary (not restricted to $\{\pm e_i\}$)
- Acceptance uses *sufficient decrease* instead of simple descent

Step-size convergence under sufficient decrease

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition $\{0, 1, 2, \dots\}$ into two index sets:

$$\mathcal{K}_s = \{k : \text{iteration } k \text{ is successful}\}, \qquad \mathcal{K}_u = \{k : \text{iteration } k \text{ is unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ **is finite.** There exists $\bar{k}$ such that all iterations $k \geq \bar{k}$ are unsuccessful, so

$$\alpha_{k+1} = \theta\alpha_k, \qquad \forall\, k \geq \bar{k}.$$

Hence $\alpha_k = \theta^{k-\bar{k}}\alpha_{\bar{k}} \to 0$. ✓

Case 2: $\mathcal{K}_s$ **is infinite.** At every $k \in \mathcal{K}_s$, sufficient decrease gives

$$f(x_k) - f(x_{k+1}) \geq \gamma\alpha_k^2.$$

Summing over the first N successful iterations $k_1 < k_2 < \cdots < k_N$ (all in $\mathcal{K}_s$):

$$f(x_0) - f(x_{k_N+1}) \geq \gamma \sum_{j=1}^{N} \alpha_{k_j}^2.$$

Step-size convergence under sufficient decrease

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition $\{0, 1, 2, \dots\}$ into two index sets:

$$\mathcal{K}_s = \{k : \text{iteration } k \text{ is successful}\}, \qquad \mathcal{K}_u = \{k : \text{iteration } k \text{ is unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ is finite. There exists $\bar{k}$ such that all iterations $k \geq \bar{k}$ are unsuccessful, so

$$\alpha_{k+1} = \theta\alpha_k, \qquad \forall\, k \geq \bar{k}.$$

Hence $\alpha_k = \theta^{k-\bar{k}}\alpha_{\bar{k}} \to 0$. ✓

Case 2: $\mathcal{K}_s$ is infinite. At every $k \in \mathcal{K}_s$, sufficient decrease gives

$$f(x_k) - f(x_{k+1}) \geq \gamma\alpha_k^2.$$

Summing over the first N successful iterations $k_1 < k_2 < \cdots < k_N$ (all in $\mathcal{K}_s$):

$$f(x_0) - f(x_{k_N+1}) \geq \gamma \sum_{j=1}^{N} \alpha_{k_j}^2.$$

Step-size convergence under sufficient decrease

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition $\{0, 1, 2, \ldots\}$ into two index sets:

$$\mathcal{K}_s = \{k : \text{iteration } k \text{ is successful}\}, \qquad \mathcal{K}_u = \{k : \text{iteration } k \text{ is unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ is finite. There exists $\bar{k}$ such that all iterations $k \geq \bar{k}$ are unsuccessful, so

$$\alpha_{k+1} = \theta \alpha_k, \qquad \forall\, k \geq \bar{k}.$$

Hence $\alpha_k = \theta^{k-\bar{k}} \alpha_{\bar{k}} \to 0$. ✓

Case 2: $\mathcal{K}_s$ is infinite. At every $k \in \mathcal{K}_s$, sufficient decrease gives

$$f(x_k) - f(x_{k+1}) \geq \gamma \alpha_k^2.$$

Summing over the first N successful iterations $k_1 < k_2 < \cdots < k_N$ (all in $\mathcal{K}_s$):

$$f(x_0) - f(x_{k_N+1}) \geq \gamma \sum_{j=1}^{N} \alpha_{k_j}^2.$$

Step-size convergence (continued)

Case 2 (continued).

By (A1), $\{x_k\} \subset L(x_0)$ and f is bounded below on $L(x_0)$ by some $f^* > -\infty$. Letting $N \to \infty$:

$$\sum_{k \in \mathcal{K}_s} \alpha_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$$

In particular, $\alpha_k^2 \to 0$ along $\mathcal{K}_s$, i.e.,

$$\lim_{k \to \infty,\, k \in \mathcal{K}_s} \alpha_k = 0.\checkmark$$

Since $\{\alpha_k\}$ is a positive non-increasing sequence, the existence of a subsequence converging to 0 implies the whole sequence converges to 0:

$$\lim_{k \to \infty} \alpha_k = 0.\square$$

Step-size convergence (continued)

Case 2 (continued).

By (A1), $\{x_k\} \subset L(x_0)$ and f is bounded below on $L(x_0)$ by some $f^* > -\infty$. Letting $N \to \infty$:

$$\sum_{k \in \mathcal{K}_s} \alpha_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$$

In particular, $\alpha_k^2 \to 0$ along $\mathcal{K}_s$, i.e.,

$$\lim_{k \to \infty,\, k \in \mathcal{K}_s} \alpha_k = 0. \checkmark$$

Since $\{\alpha_k\}$ is a positive non-increasing sequence, the existence of a subsequence converging to 0 implies the whole sequence converges to 0:

$$\lim_{k \to \infty} \alpha_k = 0. \square$$

Step-size convergence (continued)

Case 2 (continued).

By (A1), $\{x_k\} \subset L(x_0)$ and f is bounded below on $L(x_0)$ by some $f^* > -\infty$. Letting $N \to \infty$:

$$\sum_{k \in \mathcal{K}_s} \alpha_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$$

In particular, $\alpha_k^2 \to 0$ along $\mathcal{K}_s$, i.e.,

$$\lim_{k \to \infty,\, k \in \mathcal{K}_s} \alpha_k = 0. \checkmark$$

Since $\{\alpha_k\}$ is a positive non-increasing sequence, the existence of a subsequence converging to 0 implies the whole sequence converges to 0:

$$\lim_{k \to \infty} \alpha_k = 0. \square$$

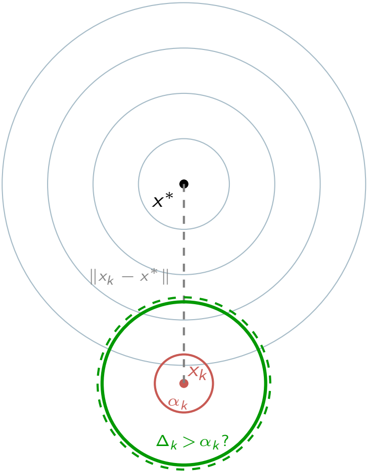
Direct Search

What else could be modified within the algorithm?

Input: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\alpha_{\min} > 0$, $\theta \in (0, 1)$, $\gamma > 0$

```
k ← 0
while α_k ≥ α_min do
    k ← k + 1
    Generate a set of poll directions D_k
    if d ∈ D_k exists such that f(x_k + α_k d) ≤ f(x_k) − γ α_k^2 then
        x_{k+1} ← x_k + α_k d, α_{k+1} ← α_k
    else
        x_{k+1} ← x_k, α_{k+1} ← θ α_k
    end if
end while
return x_k
```

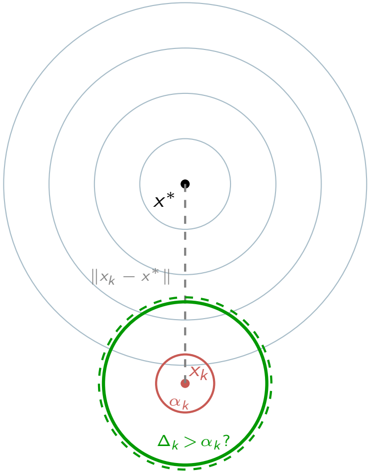
Why expand the step-size?



Observation: if the last iteration was successful with step α_k, the algorithm found a point with sufficient decrease within a ball of radius α_k around x_k.

Question: if x_k is still far from x^*, is it sensible to keep searching with the same α_k?

Why expand the step-size?



Observation: if the last iteration was successful with step α_k, the algorithm found a point with sufficient decrease within a ball of radius α_k around x_k.

Question: if x_k is still far from x^*, is it sensible to keep searching with the same α_k?

Intuition: the landscape around x_k is still "sloped" — larger steps are likely to still give sufficient decrease and would make faster progress toward x^*.

Restricting to α_k when $\|x_k - x^*\| \gg \alpha_k$ is *unnecessarily conservative*.

Idea: at successful iterations, try a *larger* step $\Delta_k \geq \alpha_k$ and accept it if sufficient decrease is still satisfied.

Direct search with step expansion

Input: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\alpha_{\min} > 0$, $\theta \in (0,1)$, $\gamma > 0$

$k \leftarrow 0$
while $\alpha_k \geq \alpha_{\min}$ **do**
    $k \leftarrow k + 1$
    Generate a set of poll directions $\mathcal{D}_k$
    if $d \in \mathcal{D}_k$ and $\Delta_k \geq \alpha_k$ exist such that
        $f(x_k + \Delta_k d) \leq f(x_k) - \gamma \Delta_k^2$ **then**
        $x_{k+1} \leftarrow x_k + \alpha_k d$, $\alpha_{k+1} \leftarrow \Delta_k$
    else
        $x_{k+1} \leftarrow x_k$, $\alpha_{k+1} \leftarrow \theta \alpha_k$
    end if
end while
return x_k

Direct search with step expansion

Input: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\alpha_{\min} > 0$, $\theta \in (0,1)$, $\gamma > 0$

$k \leftarrow 0$
while $\alpha_k \geq \alpha_{\min}$ **do**
  $k \leftarrow k + 1$
  Generate a set of poll directions $\mathcal{D}_k$
  if $d \in \mathcal{D}_k$ and $\Delta_k \geq \alpha_k$ exist such that
    $f(x_k + \Delta_k d) \leq f(x_k) - \gamma \Delta_k^2$ **then**
    $x_{k+1} \leftarrow x_k + \alpha_k d$, $\alpha_{k+1} \leftarrow \Delta_k$
  else
    $x_{k+1} \leftarrow x_k$, $\alpha_{k+1} \leftarrow \theta \alpha_k$
  end if
end while
return x_k

Key differences from the non-expansive version:

- $\{\alpha_k\}$ is *no longer non-increasing*: successful iterations may increase α_k
- The step-size used at iteration k is $\Delta_k \geq \alpha_k$, not necessarily α_k
- At failure: $\alpha_{k+1} = \theta \alpha_k < \alpha_k$ as before

Step-size convergence with expansion

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition the iterations into:

$$\mathcal{K}_s = \{k : \text{successful}\}, \qquad \mathcal{K}_u = \{k : \text{unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ is finite. Identical to before: eventually all steps are failures, giving geometric decay $\alpha_k \to 0$. ✓

Case 2: $\mathcal{K}_s$ is infinite. At every $k \in \mathcal{K}_s$, sufficient decrease with step $\Delta_k \geq \alpha_k$ gives:

$$f(x_k) - f(x_{k+1}) \geq \gamma \Delta_k^2 \geq \gamma \alpha_k^2.$$

Summing over all $k \in \mathcal{K}_s$: $$\sum_{k \in \mathcal{K}_s} \Delta_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$$

Hence $\Delta_k \to 0$ along $\mathcal{K}_s$, and since $\alpha_k \leq \Delta_k$ for $k \in \mathcal{K}_s$:

$$\lim_{k\to\infty,\, k \in \mathcal{K}_s} \alpha_k = 0. \checkmark$$

Step-size convergence with expansion

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition the iterations into:

$$\mathcal{K}_s = \{k : \text{successful}\}, \qquad \mathcal{K}_u = \{k : \text{unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ is finite. Identical to before: eventually all steps are failures, giving geometric decay $\alpha_k \to 0$. ✓

Case 2: $\mathcal{K}_s$ is infinite. At every $k \in \mathcal{K}_s$, sufficient decrease with step $\Delta_k \geq \alpha_k$ gives:

$$f(x_k) - f(x_{k+1}) \geq \gamma \Delta_k^2 \geq \gamma \alpha_k^2.$$

Summing over all $k \in \mathcal{K}_s$: $\displaystyle\sum_{k\in\mathcal{K}_s} \Delta_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$

Hence $\Delta_k \to 0$ along $\mathcal{K}_s$, and since $\alpha_k \leq \Delta_k$ for $k \in \mathcal{K}_s$:

$$\lim_{k\to\infty,\, k\in\mathcal{K}_s} \alpha_k = 0. ✓$$

Step-size convergence with expansion

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition the iterations into:

$$\mathcal{K}_s = \{k : \text{successful}\}, \qquad \mathcal{K}_u = \{k : \text{unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ is finite. Identical to before: eventually all steps are failures, giving geometric decay $\alpha_k \to 0$. ✓

Case 2: $\mathcal{K}_s$ is infinite. At every $k \in \mathcal{K}_s$, sufficient decrease with step $\Delta_k \geq \alpha_k$ gives:

$$f(x_k) - f(x_{k+1}) \geq \gamma \Delta_k^2 \geq \gamma \alpha_k^2.$$

Summing over all $k \in \mathcal{K}_s$: $\displaystyle\sum_{k\in\mathcal{K}_s} \Delta_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$

Hence $\Delta_k \to 0$ along $\mathcal{K}_s$, and since $\alpha_k \leq \Delta_k$ for $k \in \mathcal{K}_s$:

$$\lim_{k\to\infty,\, k\in\mathcal{K}_s} \alpha_k = 0. \checkmark$$

Step-size convergence with expansion

Theorem

Assume (A1). Then $\lim_{k\to\infty} \alpha_k = 0$.

Proof. Partition the iterations into:

$$\mathcal{K}_s = \{k : \text{successful}\}, \qquad \mathcal{K}_u = \{k : \text{unsuccessful}\}.$$

Case 1: $\mathcal{K}_s$ is finite. Identical to before: eventually all steps are failures, giving geometric decay $\alpha_k \to 0$. ✓

Case 2: $\mathcal{K}_s$ is infinite. At every $k \in \mathcal{K}_s$, sufficient decrease with step $\Delta_k \geq \alpha_k$ gives:

$$f(x_k) - f(x_{k+1}) \geq \gamma\Delta_k^2 \geq \gamma\alpha_k^2.$$

Summing over all $k \in \mathcal{K}_s$: $$\sum_{k\in\mathcal{K}_s} \Delta_k^2 \leq \frac{f(x_0) - f^*}{\gamma} < +\infty.$$

Hence $\Delta_k \to 0$ along $\mathcal{K}_s$, and since $\alpha_k \leq \Delta_k$ for $k \in \mathcal{K}_s$:

$$\lim_{k\to\infty,\, k\in\mathcal{K}_s} \alpha_k = 0. \checkmark$$

Step-size convergence with expansion (continued)

Case 2 (continued): connecting $\mathcal{K}_u$ to $\mathcal{K}_s$.

Since $\{\alpha_k\}$ can now *increase*, the non-increasing argument no longer applies directly. We need to track where each α_k for $k \in \mathcal{K}_u$ comes from.

Key observation: α_k can only increase at a successful iteration. For any $k \in \mathcal{K}_u$, trace back to the *last successful iteration* before k:

$$k' = \max\{j \in \mathcal{K}_s : j < k\}.$$

Between k' and k, all iterations $k'+1, \ldots, k-1$ are unsuccessful, so:

$$\alpha_k = \theta^{k-k'-1}\alpha_{k'+1} = \theta^{k-k'-1}\Delta_{k'}.$$

Since $\theta \in (0,1)$ and $k - k' \geq 1$:

$$\alpha_k \leq \Delta_{k'}.$$

As $k \to \infty$ with $k \in \mathcal{K}_u$, the index $k' \to \infty$ along $\mathcal{K}_s$ (since $\mathcal{K}_s$ is infinite), hence $\Delta_{k'} \to 0$, and therefore:

$$\lim_{k\to\infty,\, k\in\mathcal{K}_u} \alpha_k = 0. \checkmark$$

Step-size convergence with expansion (continued)

Case 2 (continued): connecting $\mathcal{K}_u$ to $\mathcal{K}_s$.

Since $\{\alpha_k\}$ can now *increase*, the non-increasing argument no longer applies directly. We need to track where each α_k for $k \in \mathcal{K}_u$ comes from.

Key observation: α_k can only increase at a successful iteration. For any $k \in \mathcal{K}_u$, trace back to the *last successful iteration* before k:

$$k' = \max\{j \in \mathcal{K}_s : j < k\}.$$

Between k' and k, all iterations $k'+1, \ldots, k-1$ are unsuccessful, so:

$$\alpha_k = \theta^{k-k'-1}\alpha_{k'+1} = \theta^{k-k'-1}\Delta_{k'}.$$

Since $\theta \in (0,1)$ and $k - k' \geq 1$:

$$\alpha_k \leq \Delta_{k'}.$$

As $k \to \infty$ with $k \in \mathcal{K}_u$, the index $k' \to \infty$ along $\mathcal{K}_s$ (since $\mathcal{K}_s$ is infinite), hence $\Delta_{k'} \to 0$, and therefore:

$$\lim_{k\to\infty,\, k\in\mathcal{K}_u} \alpha_k = 0. \checkmark$$

Step-size convergence with expansion (continued)

Case 2 (continued): connecting $\mathcal{K}_u$ to $\mathcal{K}_s$.

Since $\{\alpha_k\}$ can now *increase*, the non-increasing argument no longer applies directly. We need to track where each α_k for $k \in \mathcal{K}_u$ comes from.

Key observation: α_k can only increase at a successful iteration. For any $k \in \mathcal{K}_u$, trace back to the *last successful iteration* before k:

$$k' = \max\{j \in \mathcal{K}_s : j < k\}.$$

Between k' and k, all iterations $k' + 1, \ldots, k - 1$ are unsuccessful, so:

$$\alpha_k = \theta^{k-k'-1}\alpha_{k'+1} = \theta^{k-k'-1}\Delta_{k'}.$$

Since $\theta \in (0, 1)$ and $k - k' \geq 1$:

$$\alpha_k \leq \Delta_{k'}.$$

As $k \to \infty$ with $k \in \mathcal{K}_u$, the index $k' \to \infty$ along $\mathcal{K}_s$ (since $\mathcal{K}_s$ is infinite), hence $\Delta_{k'} \to 0$, and therefore:

$$\lim_{k\to\infty,\, k\in\mathcal{K}_u} \alpha_k = 0. \checkmark$$

Step-size convergence with expansion (continued)

Case 2 (continued): connecting $\mathcal{K}_u$ to $\mathcal{K}_s$.

Since $\{\alpha_k\}$ can now *increase*, the non-increasing argument no longer applies directly. We need to track where each α_k for $k \in \mathcal{K}_u$ comes from.

Key observation: α_k can only increase at a successful iteration. For any $k \in \mathcal{K}_u$, trace back to the *last successful iteration* before k:

$$k' = \max\{j \in \mathcal{K}_s : j < k\}.$$

Between k' and k, all iterations $k'+1, \ldots, k-1$ are unsuccessful, so:

$$\alpha_k = \theta^{k-k'-1} \alpha_{k'+1} = \theta^{k-k'-1} \Delta_{k'}.$$

Since $\theta \in (0,1)$ and $k - k' \geq 1$:

$$\alpha_k \leq \Delta_{k'}.$$

As $k \to \infty$ with $k \in \mathcal{K}_u$, the index $k' \to \infty$ along $\mathcal{K}_s$ (since $\mathcal{K}_s$ is infinite), hence $\Delta_{k'} \to 0$, and therefore:

$$\lim_{k \to \infty,\, k \in \mathcal{K}_u} \alpha_k = 0. \checkmark$$

How to compute Δ_k?

Question: We want to accept a larger step $\Delta_k \geq \alpha_k$ if it still satisfies sufficient decrease. But how do we find such Δ_k?

How to compute Δ_k?

Question: We want to accept a larger step $\Delta_k \geq \alpha_k$ if it still satisfies sufficient decrease. But how do we find such Δ_k?

Idea: Consider the *one-dimensional problem* along direction d:

$$\varphi(\alpha) = f(x_k + \alpha d), \qquad \alpha \geq 0.$$

The sufficient decrease condition becomes:

$$\varphi(\alpha) \leq \varphi(0) - \gamma\alpha^2.$$

How to compute Δ_k?

Question: We want to accept a larger step $\Delta_k \geq \alpha_k$ if it still satisfies sufficient decrease. But how do we find such Δ_k?

Idea: Consider the *one-dimensional problem* along direction d:

$$\varphi(\alpha) = f(x_k + \alpha d), \qquad \alpha \geq 0.$$

The sufficient decrease condition becomes:

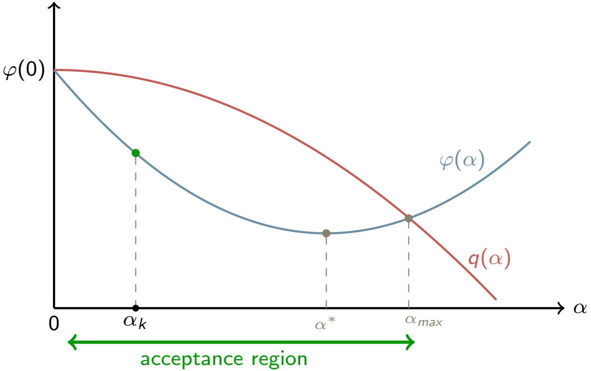
$$\varphi(\alpha) \leq \varphi(0) - \gamma\alpha^2.$$

Key observation: The right-hand side is a *quadratic function* of α:

$$q(\alpha) = \varphi(0) - \gamma\alpha^2 = f(x_k) - \gamma\alpha^2.$$

We accept step α if $\varphi(\alpha) \leq q(\alpha)$, i.e., the function stays *below the quadratic bound*.

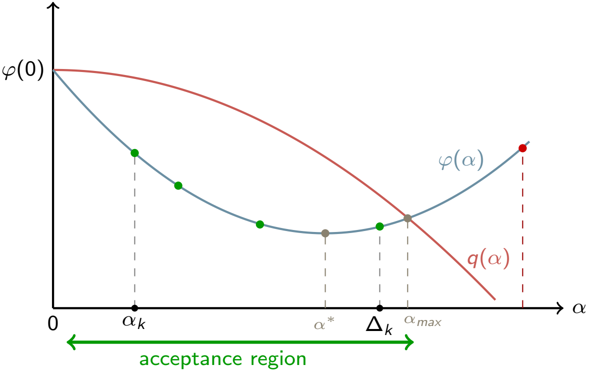
Geometric intuition: one-dimensional search



Observation: For all $\alpha \in [0, \alpha_{max}]$, we have $\varphi(\alpha) < q(\alpha)$ (function below quadratic).

Strategy: Start from α_k and *incrementally increase* α until $\varphi(\alpha) > q(\alpha)$.

Geometric intuition: one-dimensional search



Observation: For all $\alpha \in [0, \alpha_{max}]$, we have $\varphi(\alpha) < q(\alpha)$ (function below quadratic).

Strategy: Start from α_k and *incrementally increase* α until $\varphi(\alpha) > q(\alpha)$.

Computing Δ_k: successive expansion

Algorithmic approach: Try increasing multiples of α_k until sufficient decrease fails.

1: **Input:** current point x_k, direction d, step-size α_k, expansion factor $\delta > 1$
(**NOTE:** we start with $f(x_k + \alpha_k d) \leq f(x_k) - \gamma\alpha_k^2$)
2: $\alpha \leftarrow \alpha_k$
3: **while** $f(x_k + \delta\alpha \cdot d) \leq f(x_k) - \gamma(\delta\alpha)^2$ **do**
4: $\quad\alpha \leftarrow \delta\alpha$ (expand)
5: **end while**
6: **Return** $\Delta_k = \alpha$

Computing Δ_k: successive expansion

Algorithmic approach: Try increasing multiples of α_k until sufficient decrease fails.

1: **Input:** current point x_k, direction d, step-size α_k, expansion factor $\delta > 1$
(**NOTE:** we start with $f(x_k + \alpha_k d) \leq f(x_k) - \gamma\alpha_k^2$)
2: $\alpha \leftarrow \alpha_k$
3: **while** $f(x_k + \delta\alpha \cdot d) \leq f(x_k) - \gamma(\delta\alpha)^2$ **do**
4: $\quad \alpha \leftarrow \delta\alpha$ (expand)
5: **end while**
6: **Return** $\Delta_k = \alpha$

Key properties:

- The final Δ_k satisfies $\Delta_k \geq \alpha_k$ (at least one iteration with $\Delta = \alpha_k$)
- The step Δ_k satisfies sufficient decrease: $f(x_k + \Delta_k d) \leq f(x_k) - \gamma\Delta_k^2$
- Typical choice: $\rho = 2$ (doubling strategy)
- Terminates in finite iterations (function eventually violates sufficient decrease)

Why does expansion terminate?

Question: How do we know the expansion loop terminates?

Why does expansion terminate?

Question: How do we know the expansion loop terminates?

Answer: The quadratic sufficient decrease bound $q(\alpha) = f(x_k) - \gamma\alpha^2$ goes to $-\infty$ as $\alpha \to \infty$.

If f is bounded below (ensured by compactness of level sets, Assumption A1), then for large enough α:

$$\varphi(\alpha) = f(x_k + \alpha d) \geq f_{\min} > -\infty,$$

but

$$q(\alpha) = f(x_k) - \gamma\alpha^2 \to -\infty.$$

Therefore, there must exist $\bar{\alpha}$ such that $\varphi(\bar{\alpha}) > q(\bar{\alpha})$, and the loop terminates.

Direct search with step expansion

Input: $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\alpha_{\min} > 0$, $\theta \in (0,1)$, $\gamma > 0$

$k \leftarrow 0$
while $\alpha_k \geq \alpha_{\min}$ **do**
  $k \leftarrow k+1$
  Generate a set of poll directions $\mathcal{D}_k$
  if $d \in \mathcal{D}_k$ and $\Delta_k \geq \alpha_k$ exist such that
    $f(x_k + \Delta_k d) \leq f(x_k) - \gamma \Delta_k^2$ **then**
    $x_{k+1} \leftarrow x_k + \alpha_k d$, $\alpha_{k+1} \leftarrow \Delta_k$
  else
    $x_{k+1} \leftarrow x_k$, $\alpha_{k+1} \leftarrow \theta \alpha_k$
  end if
end while
return x_k

Now we know a way to compute Δ_k!

We proved that $\lim_{k \to \infty} \alpha_k = 0$. What about **stationarity**?

From step-size to stationarity: what changes?

We proved $\alpha_k \to 0$. Now we want to show that $\|\nabla f(x_k)\| \to 0$.

Recall how this was done in the simple direct search (no sufficient decrease):

At an *unsuccessful* iteration $k \in \mathcal{K}_u$, the algorithm tried *all* directions $d \in \mathcal{D}_k$ and found:

$$f(x_k + \alpha_k d) \geq f(x_k), \qquad \forall\, d \in \mathcal{D}_k.$$

Using the descent lemma with Lipschitz constant L, one could then bound:

$$\nabla f(x_k)^\top d \geq -\frac{L}{2}\alpha_k \|d\|^2, \qquad \forall\, d \in \mathcal{D}_k.$$

Now with sufficient decrease: at an unsuccessful iteration, the trial was:

$$f(x_k + \alpha_k d) \geq f(x_k) - \gamma\alpha_k^2, \qquad \forall\, d \in \mathcal{D}_k.$$

Question: This is a *weaker* rejection condition than before. Does this make the gradient bound *easier* or *harder* to obtain? Will it lead to a *larger* or *smaller* bound?

From step-size to stationarity: what changes?

We proved $\alpha_k \to 0$. Now we want to show that $\|\nabla f(x_k)\| \to 0$.

Recall how this was done in the simple direct search (no sufficient decrease):

At an *unsuccessful* iteration $k \in \mathcal{K}_u$, the algorithm tried *all* directions $d \in \mathcal{D}_k$ and found:

$$f(x_k + \alpha_k d) \geq f(x_k), \qquad \forall\, d \in \mathcal{D}_k.$$

Using the descent lemma with Lipschitz constant L, one could then bound:

$$\nabla f(x_k)^\top d \geq -\frac{L}{2}\alpha_k \|d\|^2, \qquad \forall\, d \in \mathcal{D}_k.$$

Now with sufficient decrease: at an unsuccessful iteration, the trial was:

$$f(x_k + \alpha_k d) \geq f(x_k) - \gamma \alpha_k^2, \qquad \forall\, d \in \mathcal{D}_k.$$

Question: This is a *weaker* rejection condition than before. Does this make the gradient bound *easier* or *harder* to obtain? Will it lead to a *larger* or *smaller* bound?

From step-size to stationarity: what changes?

We proved $\alpha_k \to 0$. Now we want to show that $\|\nabla f(x_k)\| \to 0$.

Recall how this was done in the simple direct search (no sufficient decrease):

At an *unsuccessful* iteration $k \in \mathcal{K}_u$, the algorithm tried *all* directions $d \in \mathcal{D}_k$ and found:

$$f(x_k + \alpha_k d) \geq f(x_k), \qquad \forall\, d \in \mathcal{D}_k.$$

Using the descent lemma with Lipschitz constant L, one could then bound:

$$\nabla f(x_k)^\top d \geq -\frac{L}{2}\alpha_k \|d\|^2, \qquad \forall\, d \in \mathcal{D}_k.$$

Now with sufficient decrease: at an unsuccessful iteration, the trial was:

$$f(x_k + \alpha_k d) \geq f(x_k) - \gamma\alpha_k^2, \qquad \forall\, d \in \mathcal{D}_k.$$

Question: This is a *weaker* rejection condition than before. Does this make the gradient bound *easier* or *harder* to obtain? Will it lead to a *larger* or *smaller* bound?

The gradient bound at unsuccessful iterations

Setup. Assume ∇f is Lipschitz continuous with constant $L > 0$:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in \mathbb{R}^n.$$

By the descent lemma, **for any direction** d **with** $\|d\| = 1$ and step $\alpha > 0$:

$$f(x_k + \alpha d) \leq f(x_k) + \alpha \nabla f(x_k)^\top d + \frac{L}{2}\alpha^2.$$

The gradient bound at unsuccessful iterations

Setup. Assume ∇f is Lipschitz continuous with constant $L > 0$:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in \mathbb{R}^n.$$

By the descent lemma, **for any direction d with** $\|d\| = 1$ and step $\alpha > 0$:

$$f(x_k + \alpha d) \leq f(x_k) + \alpha \nabla f(x_k)^\top d + \frac{L}{2}\alpha^2.$$

At an unsuccessful iteration $k \in \mathcal{K}_u$: every $d \in \mathcal{D}_k$ satisfies

$$f(x_k + \alpha_k d) \geq f(x_k) - \gamma \alpha_k^2.$$

Combining with the descent lemma:

$$f(x_k) - \gamma\alpha_k^2 \;\leq\; f(x_k + \alpha_k d) \;\leq\; f(x_k) + \alpha_k \nabla f(x_k)^\top d + \frac{L}{2}\alpha_k^2.$$

ui

Subtracting $f(x_k)$ and dividing by $\alpha_k > 0$:

$$-\nabla f(x_k)^\top d \;\leq\; \frac{L}{2}\alpha_k + \gamma\alpha_k \;=\; \left(\frac{L}{2} + \gamma\right)\alpha_k, \qquad \forall\, d \in \mathcal{D}_k.$$

From the gradient bound to stationarity

Recall the cosine measure. For a positive spanning set $\mathcal{D}_k$, the cosine measure is:

$$\kappa(\mathcal{D}_k) = \min_{v \neq 0} \max_{d \in \mathcal{D}_k} \frac{v^\top d}{\|v\| \, \|d\|} > 0.$$

In particular, for all $v \in \mathbb{R}^n$, there exists $d \in \mathcal{D}_k$ such that:

$$v^\top d \geq \kappa(\mathcal{D}_k) \|v\| \|d\|.$$

Applying the bound. Setting $v = -\nabla f(x_k)$, and considering $\|d\| = 1$ for all $d \in \mathcal{D}_k$, it holds

$$\kappa(\mathcal{D}_k)\|\nabla f(x_k)\| \;\leq\; -\nabla f(x_k)^\top d \;\leq\; \left(\frac{L}{2} + \gamma\right) \alpha_k.$$

If the cosine measure is bounded below uniformly: $\kappa(\mathcal{D}_k) \geq \kappa > 0$, then:

$$\|\nabla f(x_k)\| \;\leq\; \frac{L/2 + \gamma}{\kappa} \alpha_k, \qquad \forall\, k \in \mathcal{K}_u.$$

Since $\alpha_k \to 0$, we conclude:

$$0 \leq \liminf_{k \to \infty} \|\nabla f(x_k)\| \leq \lim_{k \to \infty,\, k \in \mathcal{K}_u} \|\nabla f(x_k)\| = 0. \checkmark$$

From the gradient bound to stationarity

Recall the cosine measure. For a positive spanning set $\mathcal{D}_k$, the cosine measure is:

$$\kappa(\mathcal{D}_k) = \min_{v \neq 0} \max_{d \in \mathcal{D}_k} \frac{v^\top d}{\|v\| \, \|d\|} > 0.$$

In particular, for all $v \in \mathbb{R}^n$, there exists $d \in \mathcal{D}_k$ such that:

$$v^\top d \geq \kappa(\mathcal{D}_k)\|v\|\|d\|.$$

Applying the bound. Setting $v = -\nabla f(x_k)$, and considering $\|d\| = 1$ for all $d \in \mathcal{D}_k$, it holds

$$\kappa(\mathcal{D}_k)\|\nabla f(x_k)\| \;\leq\; -\nabla f(x_k)^\top d \;\leq\; \left(\frac{L}{2} + \gamma\right)\alpha_k.$$

If the cosine measure is bounded below uniformly: $\kappa(\mathcal{D}_k) \geq \kappa > 0$, then:

$$\|\nabla f(x_k)\| \;\leq\; \frac{L/2 + \gamma}{\kappa}\,\alpha_k, \qquad \forall\, k \in \mathcal{K}_u.$$

Since $\alpha_k \to 0$, we conclude:

$$0 \leq \liminf_{k \to \infty} \|\nabla f(x_k)\| \leq \lim_{k \to \infty,\, k \in \mathcal{K}_u} \|\nabla f(x_k)\| = 0. \checkmark$$

From the gradient bound to stationarity

Recall the cosine measure. For a positive spanning set $\mathcal{D}_k$, the cosine measure is:

$$\kappa(\mathcal{D}_k) = \min_{v \neq 0} \max_{d \in \mathcal{D}_k} \frac{v^\top d}{\|v\| \, \|d\|} > 0.$$

In particular, for all $v \in \mathbb{R}^n$, there exists $d \in \mathcal{D}_k$ such that:

$$v^\top d \geq \kappa(\mathcal{D}_k)\|v\|\|d\|.$$

Applying the bound. Setting $v = -\nabla f(x_k)$, and considering $\|d\| = 1$ for all $d \in \mathcal{D}_k$, it holds

$$\kappa(\mathcal{D}_k)\|\nabla f(x_k)\| \ \leq \ -\nabla f(x_k)^\top d \ \leq \ \left(\frac{L}{2} + \gamma\right) \alpha_k.$$

If the cosine measure is bounded below uniformly: $\kappa(\mathcal{D}_k) \geq \kappa > 0$, then:

$$\|\nabla f(x_k)\| \ \leq \ \frac{L/2 + \gamma}{\kappa} \alpha_k, \qquad \forall\, k \in \mathcal{K}_u.$$

Since $\alpha_k \to 0$, we conclude:

$$0 \leq \liminf_{k \to \infty} \|\nabla f(x_k)\| \leq \lim_{k \to \infty,\, k \in \mathcal{K}_u} \|\nabla f(x_k)\| = 0. \checkmark$$

From the gradient bound to stationarity

Recall the cosine measure. For a positive spanning set $\mathcal{D}_k$, the cosine measure is:

$$\kappa(\mathcal{D}_k) = \min_{v \neq 0} \max_{d \in \mathcal{D}_k} \frac{v^\top d}{\|v\| \, \|d\|} > 0.$$

In particular, for all $v \in \mathbb{R}^n$, there exists $d \in \mathcal{D}_k$ such that:

$$v^\top d \geq \kappa(\mathcal{D}_k) \|v\| \|d\|.$$

Applying the bound. Setting $v = -\nabla f(x_k)$, and considering $\|d\| = 1$ for all $d \in \mathcal{D}_k$, it holds

$$\kappa(\mathcal{D}_k)\|\nabla f(x_k)\| \ \leq \ -\nabla f(x_k)^\top d \ \leq \ \left(\frac{L}{2} + \gamma\right) \alpha_k.$$

If the cosine measure is bounded below uniformly: $\kappa(\mathcal{D}_k) \geq \kappa > 0$, then:

$$\|\nabla f(x_k)\| \ \leq \ \frac{L/2 + \gamma}{\kappa} \, \alpha_k, \qquad \forall \, k \in \mathcal{K}_u.$$

Since $\alpha_k \to 0$, we conclude:

$$0 \leq \liminf_{k \to \infty} \|\nabla f(x_k)\| \leq \lim_{k \to \infty, \, k \in \mathcal{K}_u} \|\nabla f(x_k)\| = 0. \checkmark$$

Stationarity: summary of the results

Assumptions:

- (A1) The level set $L(x_0) = \{x : f(x) \leq f(x_0)\}$ is compact
- (A2) ∇f is Lipschitz continuous with constant $L > 0$
- (A3) The cosine measure is uniformly bounded: $\kappa(\mathcal{D}_k) \geq \kappa > 0$ for all k

Theorem (Gradient bound at unsuccessful iterations)

Consider the direct search algorithm with sufficient decrease.
Under assumptions (A1), (A2), and (A3), at any unsuccessful iteration $k \in \mathcal{K}_u$ it holds

$$\|\nabla f(x_k)\| \leq \frac{L/2 + \gamma}{\kappa} \alpha_k.$$

Theorem (Convergence to stationary points)

Under assumptions (A1), (A2), (A3), the direct search algorithm with sufficient decrease generates a sequence $\{x_k\}$ satisfying

$$\liminf_{k \to \infty} \|\nabla f(x_k)\| = 0.$$

A practical question

So far we proved that the algorithms we studied *eventually* converge. But. . .

Question: In practice, can we run an algorithm forever?

A practical question

So far we proved that the algorithms we studied *eventually* converge. But...

Question: In practice, can we run an algorithm forever?

Of course not. Every algorithm must **stop at some point**, returning an *approximate* solution rather than the exact one.

A practical question

So far we proved that the algorithms we studied *eventually* converge. But. . .

Question: In practice, can we run an algorithm forever?

Of course not. Every algorithm must **stop at some point**, returning an *approximate* solution rather than the exact one.

This raises two natural questions:

1. How do we *measure* how good an approximate solution is?

A practical question

So far we proved that the algorithms we studied *eventually* converge. But. . .

Question: In practice, can we run an algorithm forever?

Of course not. Every algorithm must **stop at some point**, returning an *approximate* solution rather than the exact one.

This raises two natural questions:

1. How do we *measure* how good an approximate solution is?
2. How many iterations do we need to reach a desired level of approximation?

Measuring precision: three natural candidates

Measuring precision: three natural candidates

Three natural measures of precision at iteration k:

1. **Norm of the gradient:**

$$\|\nabla f(x_k)\| \leq \epsilon$$

   x_k is an *approximate stationary point* — the function is nearly flat at x_k.

Measuring precision: three natural candidates

Three natural measures of precision at iteration k:

1. **Norm of the gradient:**

$$\|\nabla f(x_k)\| \leq \epsilon$$

x_k is an *approximate stationary point* — the function is nearly flat at x_k.

2. **Distance to the optimal function value:**

$$f(x_k) - f^* \leq \epsilon$$

x_k is an *approximate global minimizer in function value* — we are nearly as good as possible.

Measuring precision: three natural candidates

Three natural measures of precision at iteration k:

1. **Norm of the gradient:**

$$\|\nabla f(x_k)\| \leq \epsilon$$

   x_k is an *approximate stationary point* — the function is nearly flat at x_k.

2. **Distance to the optimal function value:**

$$f(x_k) - f^* \leq \epsilon$$

   x_k is an *approximate global minimizer in function value* — we are nearly as good as possible.

3. **Distance to the optimal solution:**

$$\|x_k - x^*\| \leq \epsilon$$

   x_k is an *approximate global minimizer in the decision space.*

Measuring precision: three natural candidates

There is no unique answer — the right choice depends on what we know about the problem.

Three natural measures of precision at iteration k:

1. **Norm of the gradient:**

$$\|\nabla f(x_k)\| \leq \epsilon$$

   x_k is an *approximate stationary point* — the function is nearly flat at x_k.

2. **Distance to the optimal function value:**

$$f(x_k) - f^* \leq \epsilon$$

   x_k is an *approximate global minimizer in function value* — we are nearly as good as possible.

3. **Distance to the optimal solution:**

$$\|x_k - x^*\| \leq \epsilon$$

   x_k is an *approximate global minimizer in the decision space.*

Question: Can we always use all three? Are they always well-defined?

Which measure can we use?

The answer depends on the **structure of the problem**.

Problem class	$\|\nabla f(x_k)\|$	$f(x_k) - f^*$	$\|x_k - x^*\|$
General (nonconvex)	✓	×	×
Convex	✓	✓	×
Strongly convex	✓	✓	✓

Nonconvex problems: stationary points are not necessarily global minima, so f^* and x^* cannot be used as targets. The gradient norm is the only sensible measure.

Convex problems: every stationary point *is* a global minimum, so $f(x_k) - f^* \leq \epsilon$ is meaningful. However, x^* may *not be unique*.

Strongly convex problems: the global minimizer x^* is *unique*, so all three measures are well-defined.

Worst-case complexity bounds

We want to answer the following question

"How many iterations suffice to guarantee precision ϵ, in the worst case?"

Iteration complexity: Fix a target precision ϵ. How many iterations $T(\epsilon)$ do we need?

$$\text{To reach } \|\nabla f(x_k)\| \leq \epsilon : \quad T(\epsilon) = ?$$

Worst-case complexity bounds

We want to answer the following question

"How many iterations suffice to guarantee precision ϵ, in the worst case?"

Iteration complexity: Fix a target precision ϵ. How many iterations $T(\epsilon)$ do we need?

$$\text{To reach } \|\nabla f(x_k)\| \leq \epsilon : \quad T(\epsilon) = ?$$

Worst-case complexity bounds

We want to answer the following question

"How many iterations suffice to guarantee precision ϵ, in the worst case?"

Iteration complexity: Fix a target precision ϵ. How many iterations $T(\epsilon)$ do we need?

$$\text{To reach } \|\nabla f(x_k)\| \leq \epsilon : \quad T(\epsilon) = ?$$

Convergence rate

The WCC bound $T(\epsilon)$ tells us *how fast* the algorithm converges. Different rates have very different practical implications.

Rate	Sublinear	Linear	Quadratic
Iterations to reach ϵ	$\mathcal{O}(1/\epsilon^p)$, $p > 0$	$\mathcal{O}(\log(1/\epsilon))$	$\mathcal{O}(\log\log(1/\epsilon))$

Key takeaway: sublinear rates are typical for nonconvex problems (for gradient-based methods). To go from $\epsilon = 10^{-2}$ to $\epsilon = 10^{-4}$: (double precision)

- Sublinear ($p = 2$): requires $\times 100$ more iterations
- Linear: requires only $\sim 2\times$ more iterations
- Quadratic: barely any more iterations at all

Convergence rate

The WCC bound $T(\epsilon)$ tells us *how fast* the algorithm converges. Different rates have very different practical implications.

Rate	Sublinear	Linear	Quadratic
Iterations to reach ϵ	$\mathcal{O}(1/\epsilon^p)$, $p > 0$	$\mathcal{O}(\log(1/\epsilon))$	$\mathcal{O}(\log\log(1/\epsilon))$

Key takeaway: sublinear rates are typical for nonconvex problems (for gradient-based methods). To go from $\epsilon = 10^{-2}$ to $\epsilon = 10^{-4}$: (double precision)

- Sublinear ($p = 2$): requires $\times 100$ more iterations
- Linear: requires only $\sim 2\times$ more iterations
- Quadratic: barely any more iterations at all