First-Order Methods and Hyperbolic Programming

Jim Renegar

Cornell ORIE

Simons, Spring 2019

First-Order Methods (FOM's)

convex function

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in Q \end{array}$$

closed convex set

"step size"

$$x_{k+1} = P_Q(x_k - \alpha_k g_k)$$

orthogonal projection onto Q

gradient (or subgradient) of f at x_k

Unless Q is a simple set, the bottleneck is orthogonal projection.

"Proximal" methods replace orthogonal projection with a different operation, but all sets for which that operation can be done efficiently are simple.

In the context of differentiable objective functions, it is usually assumed that f is "smooth", meaning there exists a constant L satisfying $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ for all x, y in an open neighborhood of Q.

Then choosing $\alpha_k = 1/L$, and letting ϵ be a positive scalar,

$$k \geq L\,\mathrm{dist}(x_0, X^*)^2/(2\epsilon) \quad \Rightarrow \quad f(x_k) \leq f^* + \epsilon$$

set of optimal solutions

optimal objective value

First-Order Methods (FOM's)

convex function

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in Q \end{array}$$

closed convex set

"step size"

$$x_{k+1} = P_Q(x_k - \alpha_k g_k)$$

orthogonal projection onto Q

gradient (or subgradient) of f at x_k

Unless Q is a simple set, the bottleneck is orthogonal projection.

"Proximal" methods replace orthogonal projection with a different operation, but all sets for which that operation can be done efficiently are simple.

In the context of differentiable objective functions, it is usually assumed that f is "smooth", meaning there exists a constant L satisfying $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ for all x, y in an open neighborhood of Q.

For Nesterov's (first) accelerated method,

$$k \geq 2\,\mathrm{dist}(x_0, X^*)\,\sqrt{L/\epsilon} \quad \Rightarrow \quad f(x_k) \leq f^* + \epsilon$$

set of optimal solutions

optimal objective value

$$\boxed{\begin{aligned} x_{k+1} &= P_Q\big(y_k - \tfrac{1}{L}\nabla f(y_k)\big) \\ \theta_{k+1} &= \tfrac{1}{2}\,\big(1 + \sqrt{1 + 4\theta_k^2}\,\big) \\ y_{k+1} &= x_{k+1} + \tfrac{\theta_k - 1}{\theta_{k+1}}\,(x_{k+1} - x_k) \end{aligned}}$$

First-Order Methods (FOM's)

convex function

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in Q \end{array}$$

closed convex set

"step size"

$$x_{k+1} = P_Q(x_k - \alpha_k g_k)$$

orthogonal projection onto Q

gradient (or subgradient) of f at x_k

Unless Q is a simple set, the bottleneck is orthogonal projection.

"Proximal" methods replace orthogonal projection with a different operation, but all sets for which that operation can be done efficiently are simple.

In the context of non-differentiable objective functions, it is usually assumed that f is Lipschitz, that is, there exists a constant M satisfying $|f(x) - f(y)| \leq M\|x - y\|$ for all x, y in an open neighborhood of Q.

Then choosing $\alpha_k = \epsilon/\|g_k\|^2$,

$$\ell \geq \big(M \operatorname{dist}(x_0, X^*)/\epsilon\big)^2 \quad \Rightarrow \quad \min_{k \leq \ell} f(x_k) \leq f^* + \epsilon$$

set of optimal solutions

optimal objective value

First-Order Methods (FOM's)

convex function

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in Q \end{aligned}$$

closed convex set

"step size"

$$x_{k+1} = P_Q(x_k - \alpha_k g_k)$$

orthogonal projection onto Q

gradient (or subgradient) of f at x_k

Unless Q is a simple set, the bottleneck is orthogonal projection.

"Proximal" methods replace orthogonal projection with a different operation, but all sets for which that operation can be done efficiently are simple.

---

gradient method	accelerated method	subgradient method
$O\left(\frac{L}{\epsilon}\,\mathrm{dist}(x_0, X^*)^2\right)$	$O\left(\sqrt{\frac{L}{\epsilon}}\,\mathrm{dist}(x_0, X^*)\right)$	$O\left(\left(\frac{M}{\epsilon}\right)^2 \mathrm{dist}(x_0, X^*)^2\right)$

Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle \ , \ \rangle$ and norm $\| \ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\mathrm{int}(\mathcal{K})$, zero on $\mathrm{bdy}(\mathcal{K})$,
and for each $e \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

<u>Fix</u> $e \in \mathrm{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$
and centrally symmetric around e

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

$|\lambda_{\min}(x) - \lambda_{\min}(y)| \le \|x - y\|_\infty$ for all $x, y \in \mathcal{E}$

i.e., Lipschitz constant $= 1$ w.r.t. the norm $\| \ \|_\infty$

Consequently, $\lambda_{\min}(x) - \lambda_{\min}(y)| \le \frac{1}{r_e}\|x - y\|$ Euclidean norm

where r_e is the largest radius of a Euclidean ball centered at e
and contained in $\mathcal{K}$

Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\text{int}(\mathcal{K})$, zero on $\text{bdy}(\mathcal{K})$,
and for each $e \in \text{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

<u>Fix</u> $e \in \text{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$
and centrally symmetric around e

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

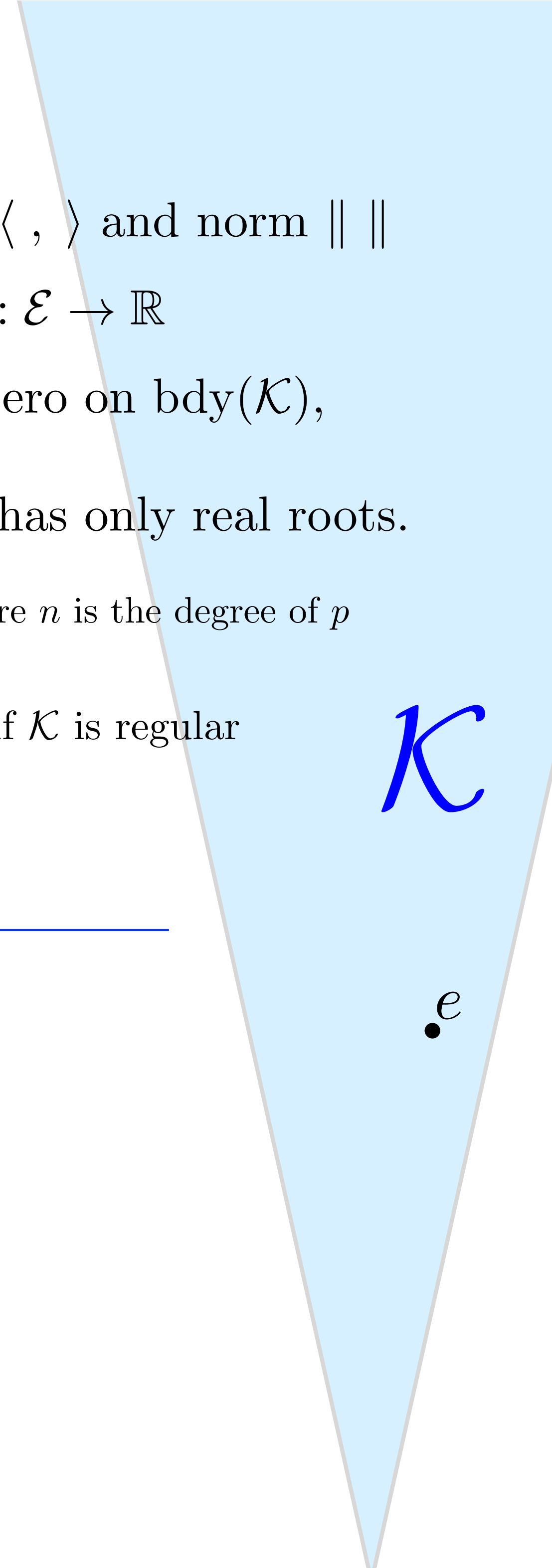
Thus, p is homogeneous, nonzero on $\text{int}(\mathcal{K})$, zero on $\text{bdy}(\mathcal{K})$,
and for each $e \in \text{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

<u>Fix</u> $e \in \text{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$
and centrally symmetric around e

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x



Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle \ , \ \rangle$ and norm $\| \ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

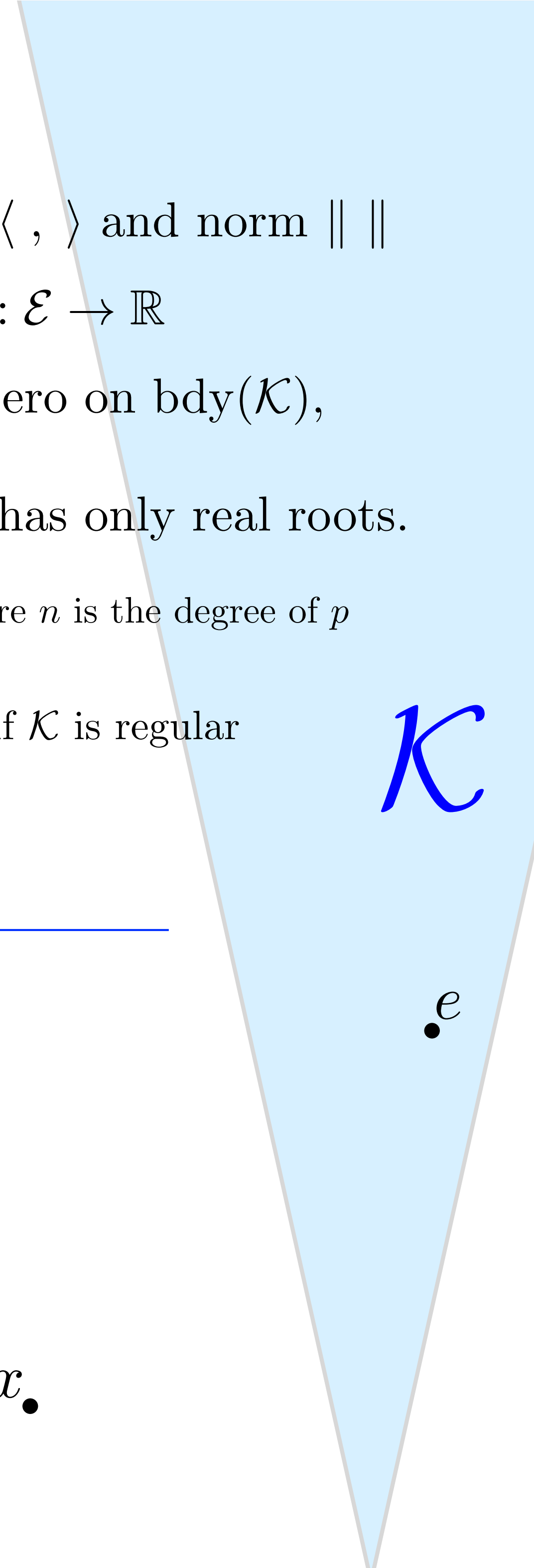
Thus, p is homogeneous, nonzero on $\mathrm{int}(\mathcal{K})$, zero on $\mathrm{bdy}(\mathcal{K})$,
and for each $e \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

<u>Fix</u> $e \in \mathrm{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$
and centrally symmetric around e

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x



Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\mathrm{int}(\mathcal{K})$, zero on $\mathrm{bdy}(\mathcal{K})$,
and for each $e \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

Fix $e \in \mathrm{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

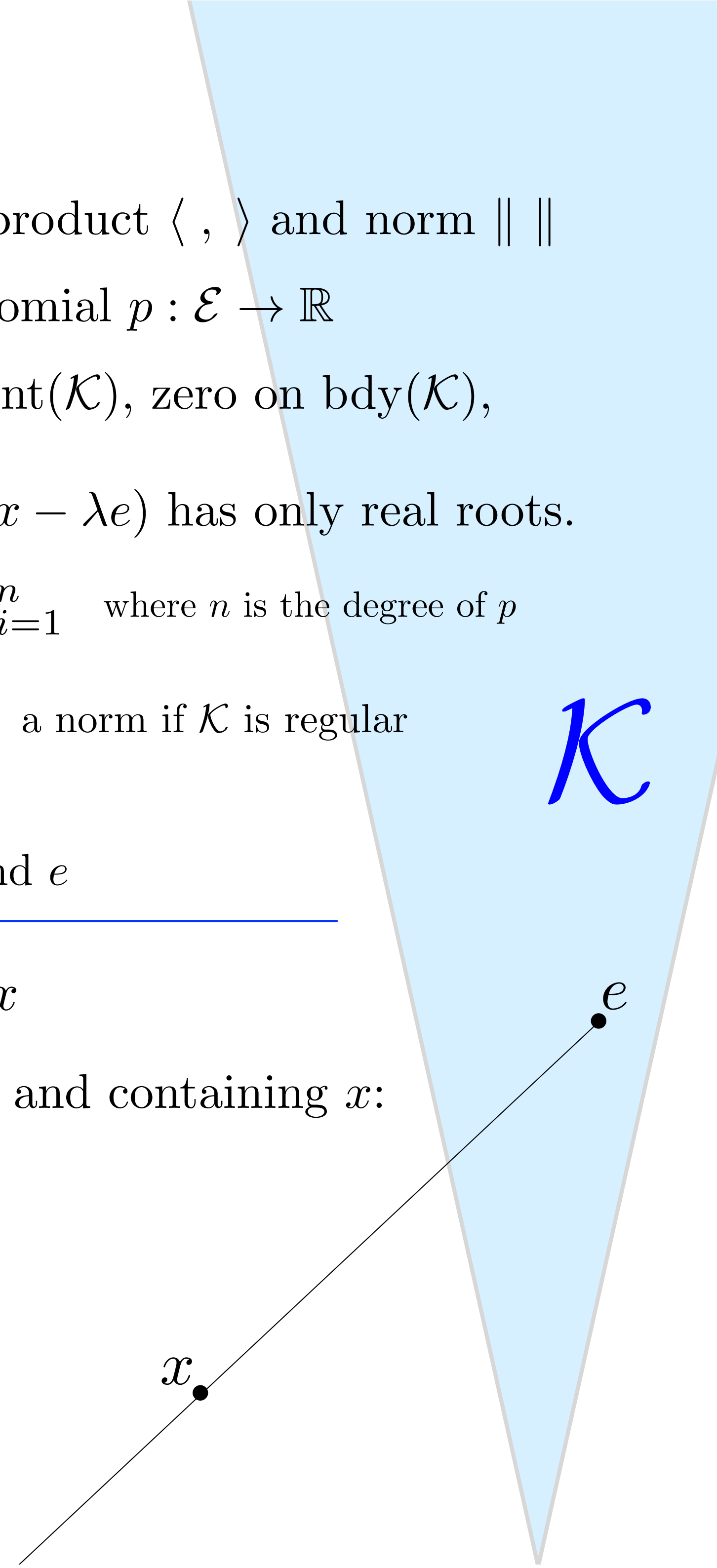
$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$ and centrally symmetric around e

---

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

For $x \in \mathcal{E}$, consider the half-line with endpoint e and containing x:
$\{e + t(x - e) : t \geq 0\}$



Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\mathrm{int}(\mathcal{K})$, zero on $\mathrm{bdy}(\mathcal{K})$,
and for each $e \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

Fix $e \in \mathrm{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$
and centrally symmetric around e

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

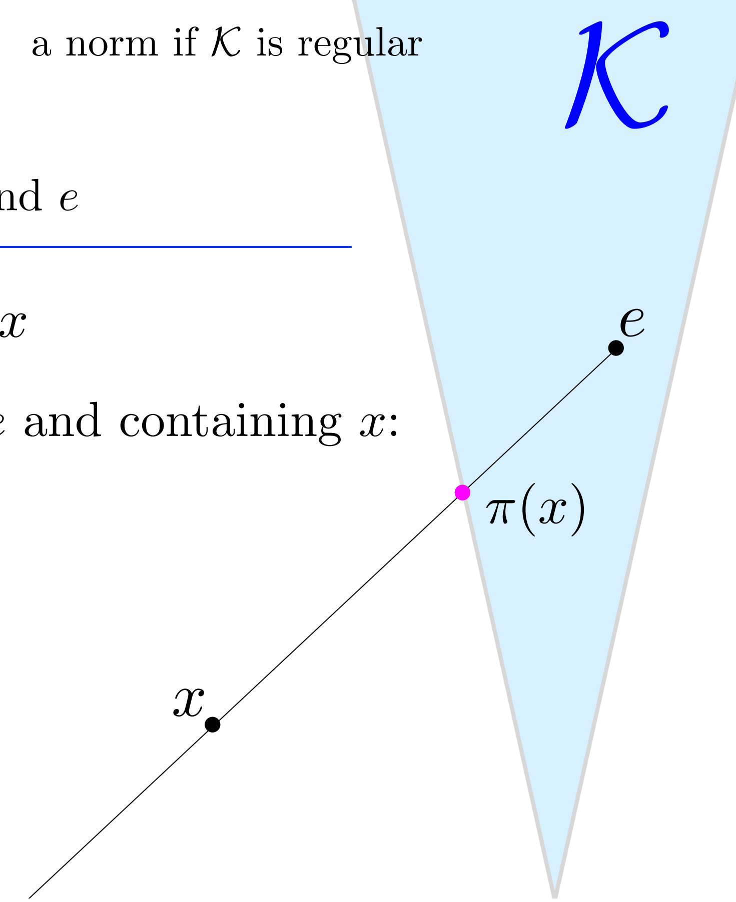
For $x \in \mathcal{E}$, consider the half-line with endpoint e and containing x:

$$\{e + t(x - e) : t \geq 0\}$$

The half-line intersects $\mathrm{bdy}(\mathcal{K})$ iff $\lambda_{\min}(x) < 1$,
in which case the point of intersection is

$$\pi(x) := e + \frac{1}{1 - \lambda_{\min}(x)}(x - e)$$

"the radial projection of x to $\mathrm{bdy}(\mathcal{K})$"



Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\text{int}(\mathcal{K})$, zero on $\text{bdy}(\mathcal{K})$,
and for each $e \in \text{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

<u>Fix</u> $e \in \text{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$ and centrally symmetric around e

---

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

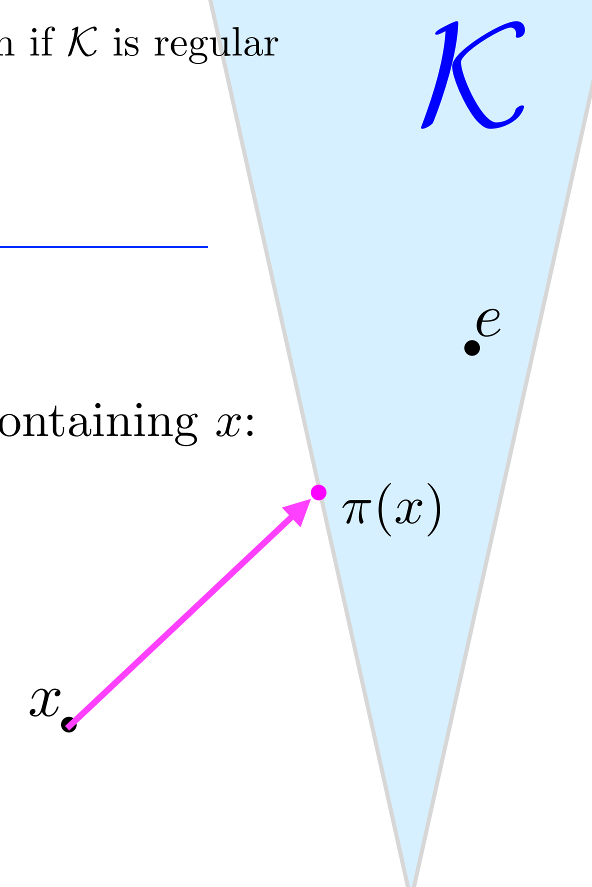
For $x \in \mathcal{E}$, consider the half-line with endpoint e and containing x:

$$\{e + t(x - e) : t \geq 0\}$$

The half-line intersects $\text{bdy}(\mathcal{K})$ iff $\lambda_{\min}(x) < 1$,
in which case the point of intersection is

$$\pi(x) := e + \tfrac{1}{1 - \lambda_{\min}(x)}(x - e)$$

"the radial projection of x to $\text{bdy}(\mathcal{K})$"



Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\mathrm{int}(\mathcal{K})$, zero on $\mathrm{bdy}(\mathcal{K})$,
and for each $e \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

Fix $e \in \mathrm{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e,1)$ largest set both contained in $\mathcal{K}$ and centrally symmetric around e

---

$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

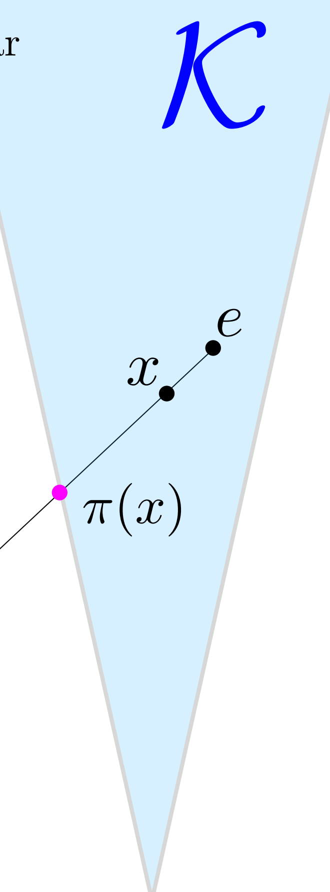
For $x \in \mathcal{E}$, consider the half-line with endpoint e and containing x:

$$\{e + t(x - e) : t \geq 0\}$$

The half-line intersects $\mathrm{bdy}(\mathcal{K})$ iff $\lambda_{\min}(x) < 1$,
in which case the point of intersection is

$$\pi(x) := e + \frac{1}{1 - \lambda_{\min}(x)}(x - e)$$

"the radial projection of x to $\mathrm{bdy}(\mathcal{K})$"



Hyperbolicity Cones

$\mathcal{E}$ finite-dimensional Euclidean space with inner product $\langle\ ,\ \rangle$ and norm $\|\ \|$

$\mathcal{K} \subset \mathcal{E}$ hyperbolicity cone with hyperbolic polynomial $p : \mathcal{E} \to \mathbb{R}$

Thus, p is homogeneous, nonzero on $\mathrm{int}(\mathcal{K})$, zero on $\mathrm{bdy}(\mathcal{K})$,
and for each $e \in \mathrm{int}(\mathcal{K})$ and $x \in \mathcal{E}$,
the univariate polynomial $\lambda \mapsto p(x - \lambda e)$ has only real roots.

<u>Fix</u> $e \in \mathrm{int}(\mathcal{K})$ and denote the roots as $\{\lambda_i(x)\}_{i=1}^n$ where n is the degree of p

$\|x\|_\infty := \max_i |\lambda_i(x)|$ a seminorm on $\mathcal{E}$ a norm if $\mathcal{K}$ is regular

closed ball of radius 1
$\bar{B}_\infty(e, 1)$ largest set both contained in $\mathcal{K}$
and centrally symmetric around e

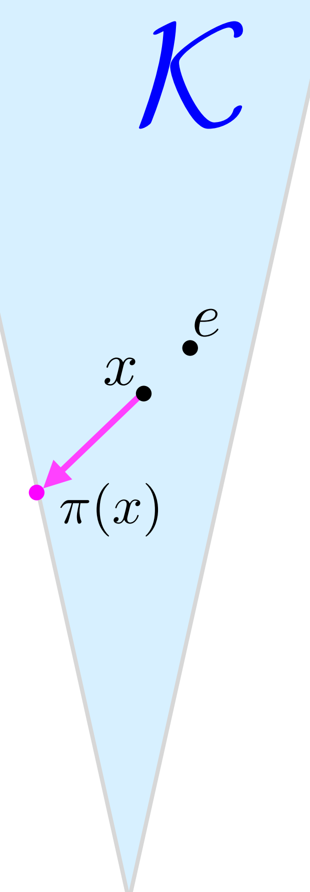
$\lambda_{\min}(x) := \min_i \lambda_i(x)$ a concave function in x

For $x \in \mathcal{E}$, consider the half-line with endpoint e and containing x:
$\{e + t(x - e) : t \geq 0\}$

The half-line intersects $\mathrm{bdy}(\mathcal{K})$ iff $\lambda_{\min}(x) < 1$,
in which case the point of intersection is

$$\pi(x) := e + \tfrac{1}{1 - \lambda_{\min}(x)}(x - e)$$

"the radial projection of x to $\mathrm{bdy}(\mathcal{K})$"



Hyperbolic Programming

$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K} \end{aligned}$$

Trivially, the objective function is smooth,
so gradient and accelerated methods can be applied:

$$x_{k+1} = P_Q(x_k - \alpha_k c)$$
where $Q = \{x \in \mathcal{K} \mid Ax = b\}$

For high-dimensional problems,
the orthogonal projection is impractical except in special cases.

Are there practical ways of applying FOM's to solve hyperbolic programs?

---

the point we fixed in $\text{int}(\mathcal{K})$

Assume e satisfies $Ae = b$

In other words,
assume we know
a strictly feasible point e

Assume the hyperbolic program has an optimal solution.

Fix a scalar z satisfying $z < \langle c, e \rangle$ and consider the optimization problem

$$\begin{aligned} \max \quad & \lambda_{\min}(x) \\ \text{s.t.} \quad & Ax = b \\ & \langle c, x \rangle = z \end{aligned}$$

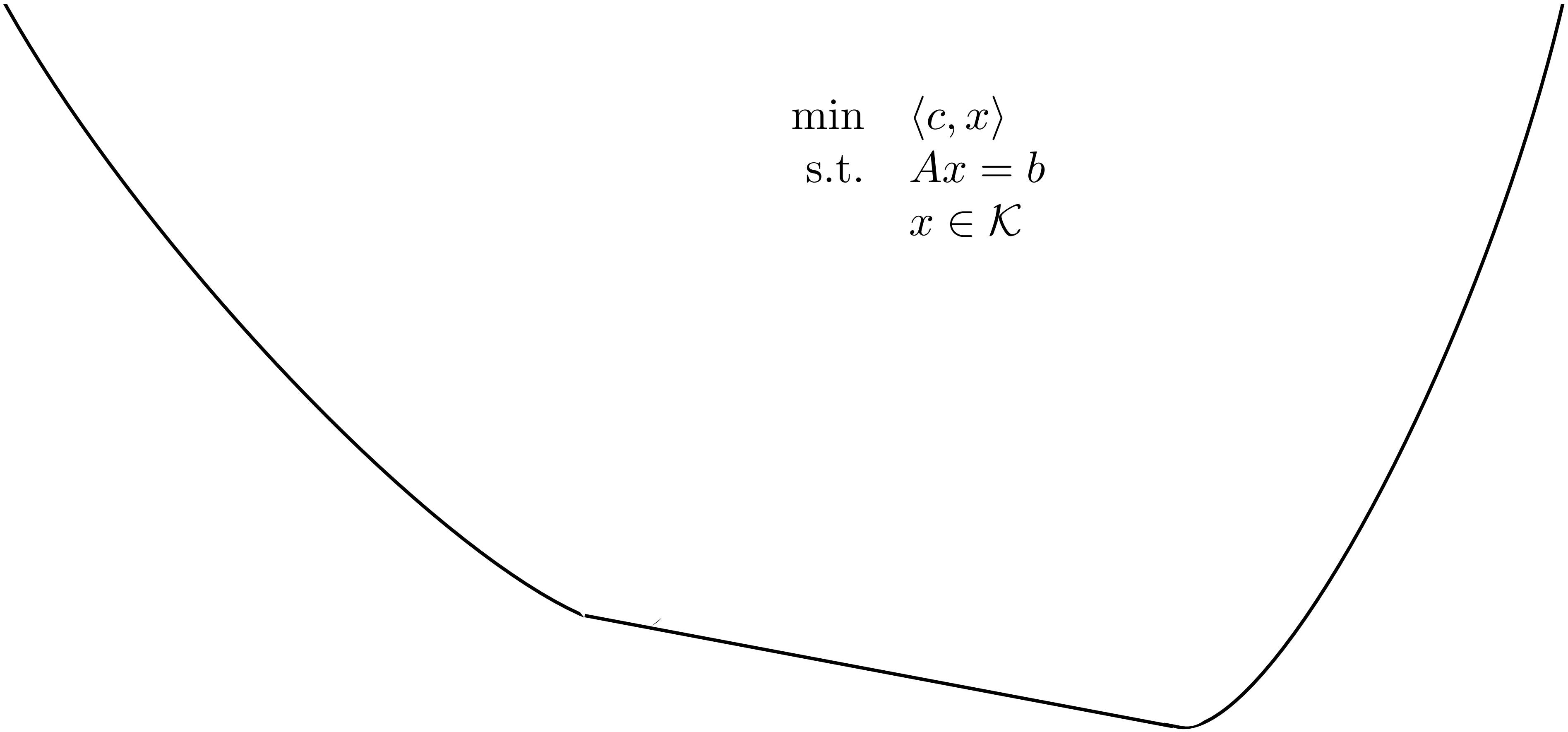
Here the feasible region is an affine space,
and so computing orthogonal projections is easy

Claim: A point x is optimal for the eigenvalue optimization problem
if and only if $\pi(x)$ is optimal for the hyperbolic program.

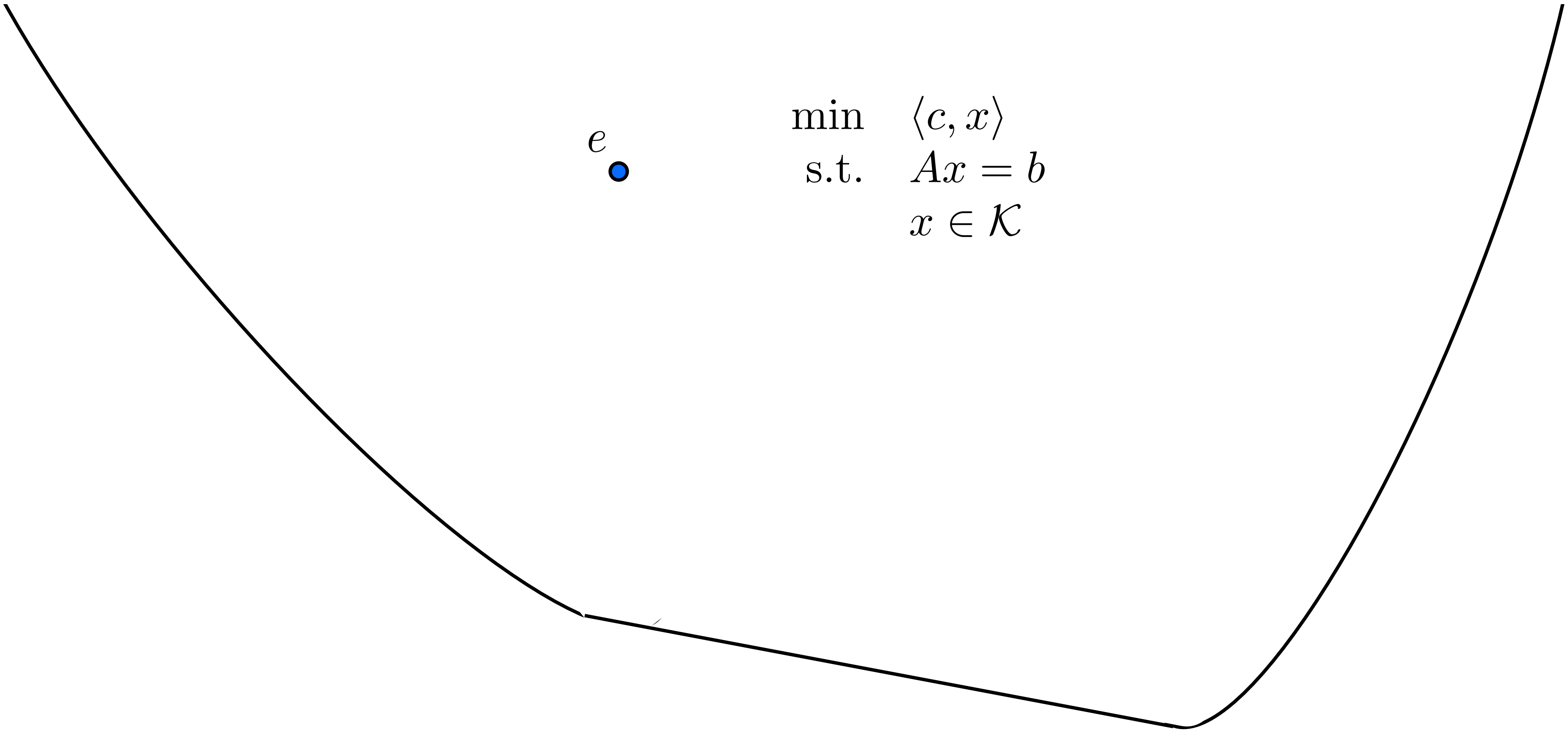
$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K} \end{aligned}$$

Think of this 2-dimensional plane
as being the slice of $\mathcal{E}$
cut out by $\{x \mid Ax = b\}$

$$
\begin{aligned}
\min \quad & \langle c, x \rangle \\
\text{s.t.} \quad & Ax = b \\
& x \in \mathcal{K}
\end{aligned}
$$

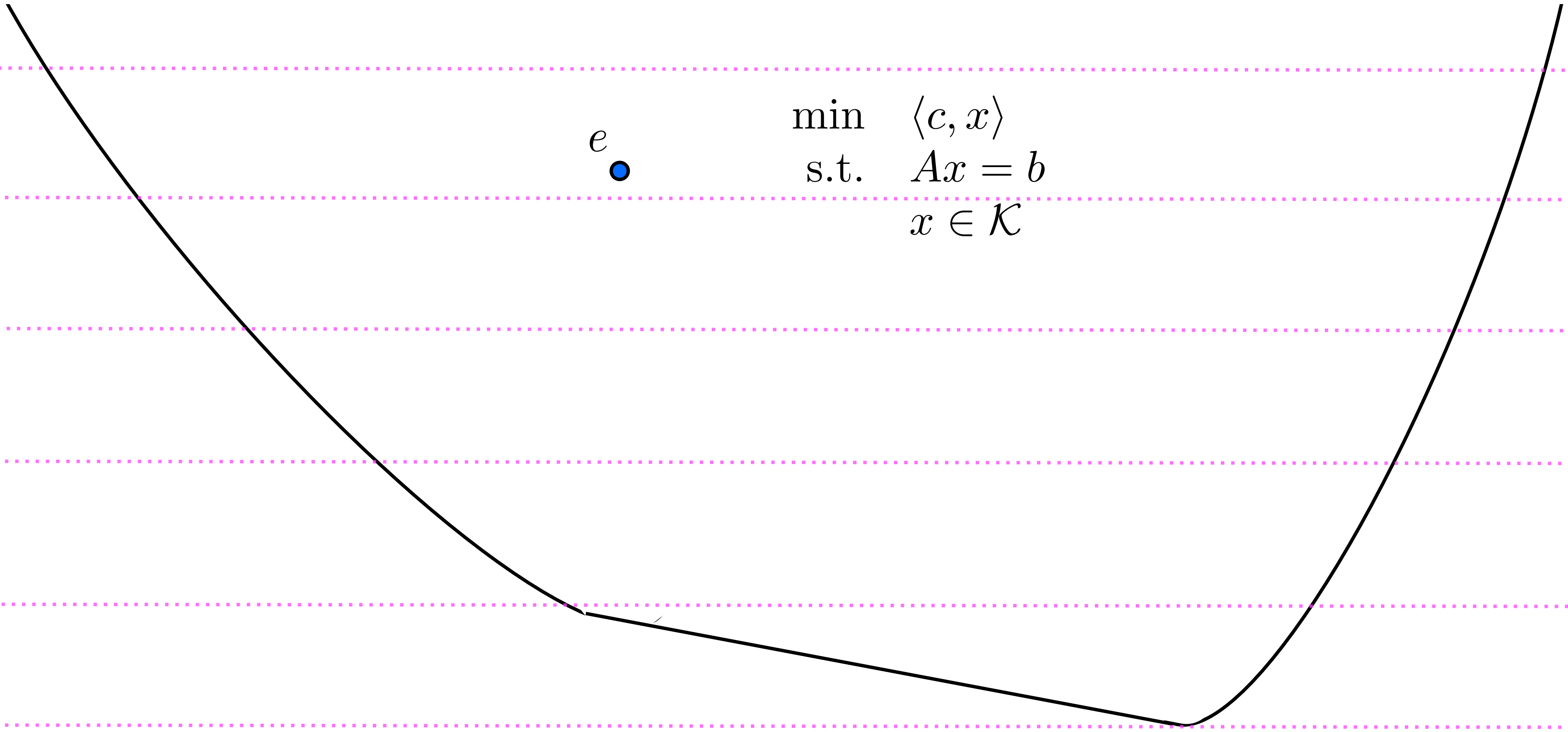


Think of this 2-dimensional plane
as being the slice of $\mathcal{E}$
cut out by $\{x \mid Ax = b\}$



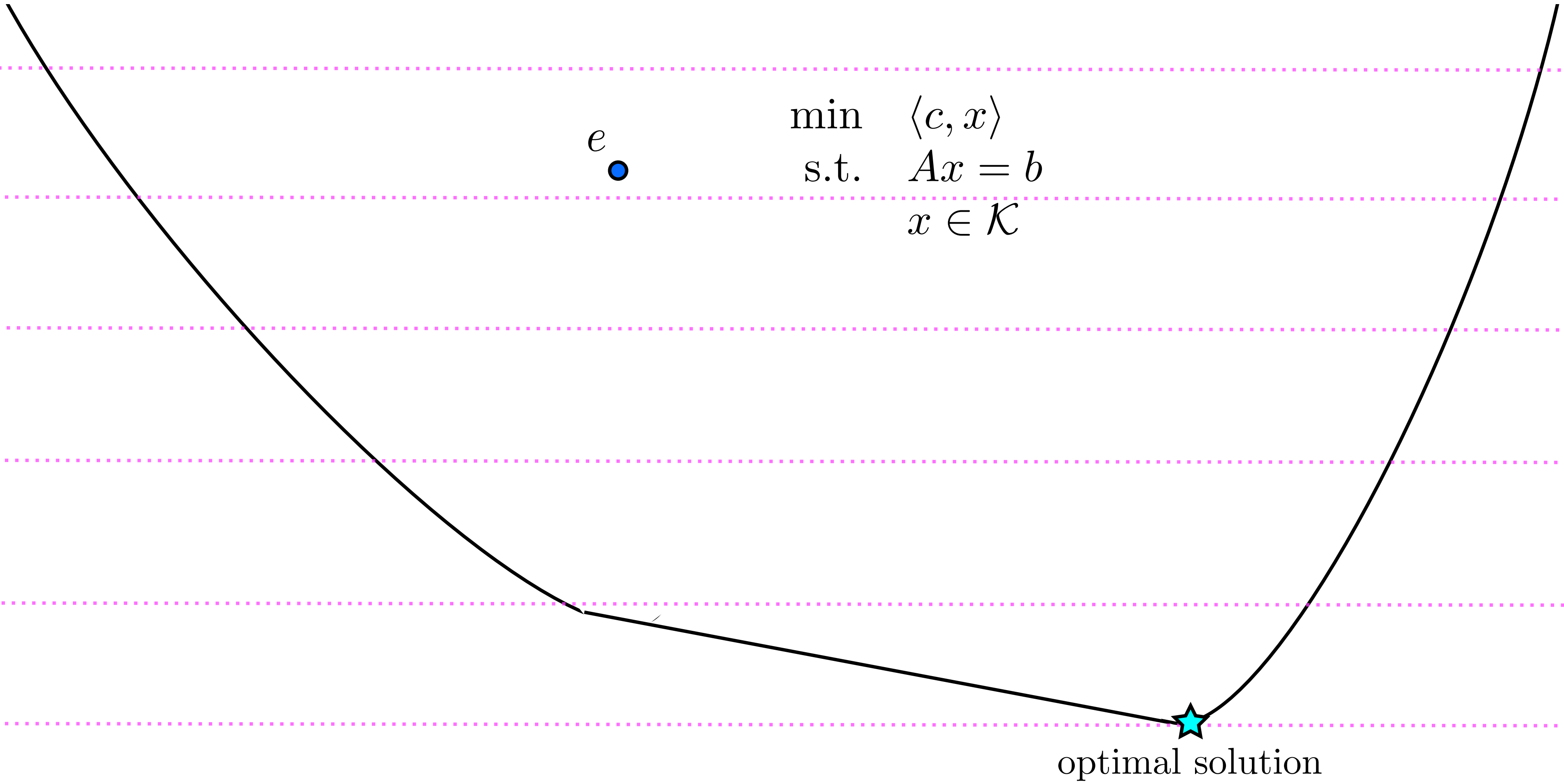
$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K} \end{aligned}$$

Think of this 2-dimensional plane
as being the slice of $\mathcal{E}$
cut out by $\{x \mid Ax = b\}$



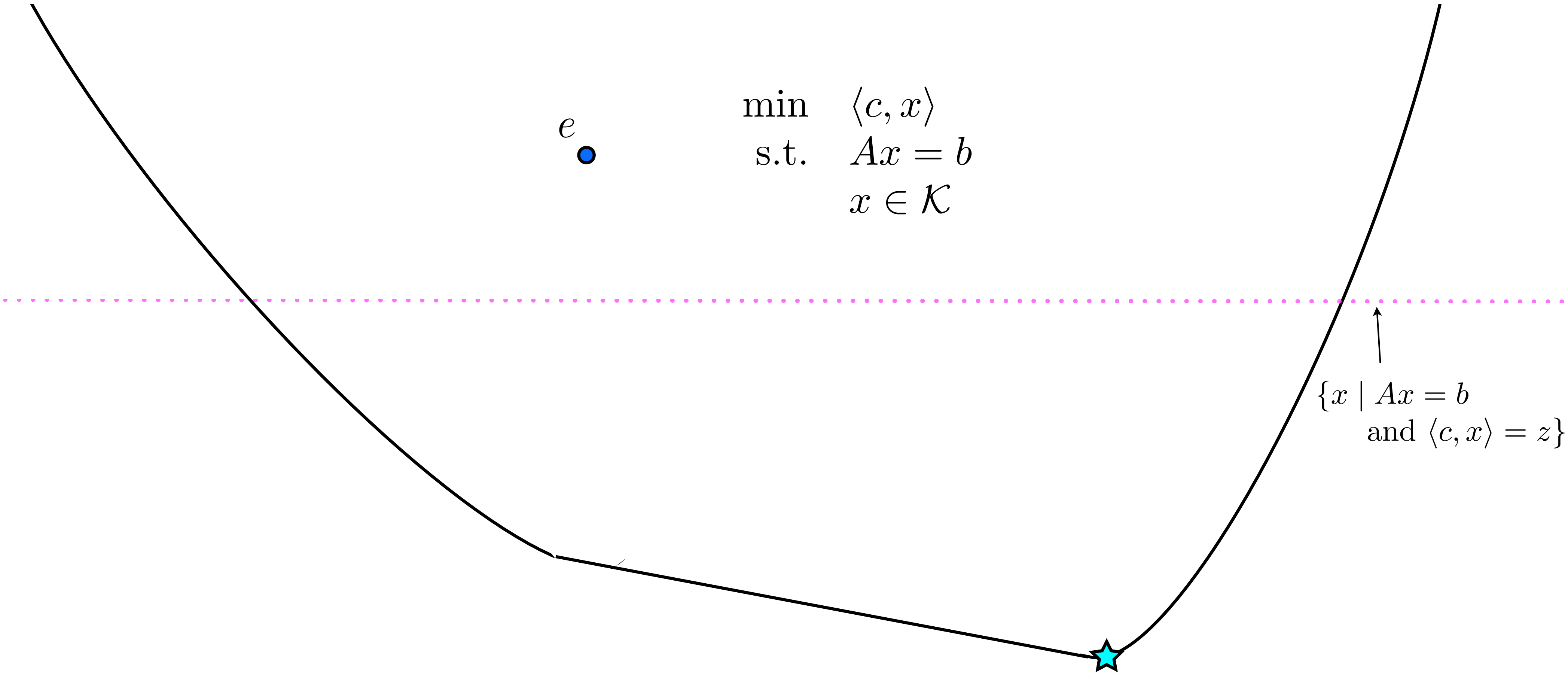
$$\begin{array}{rl} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}$$

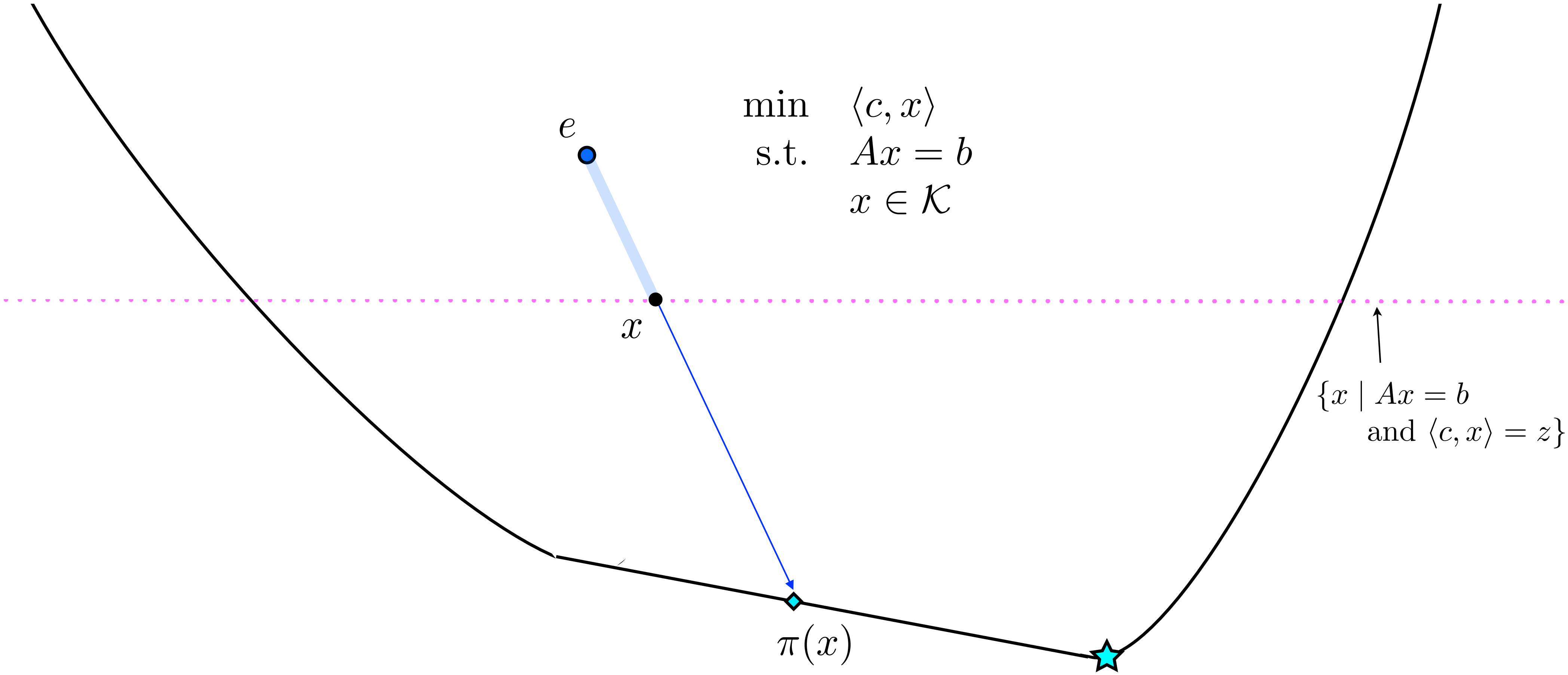
Assume the objective function $x \mapsto \langle c, x\rangle$ is constant on horizontal slices.



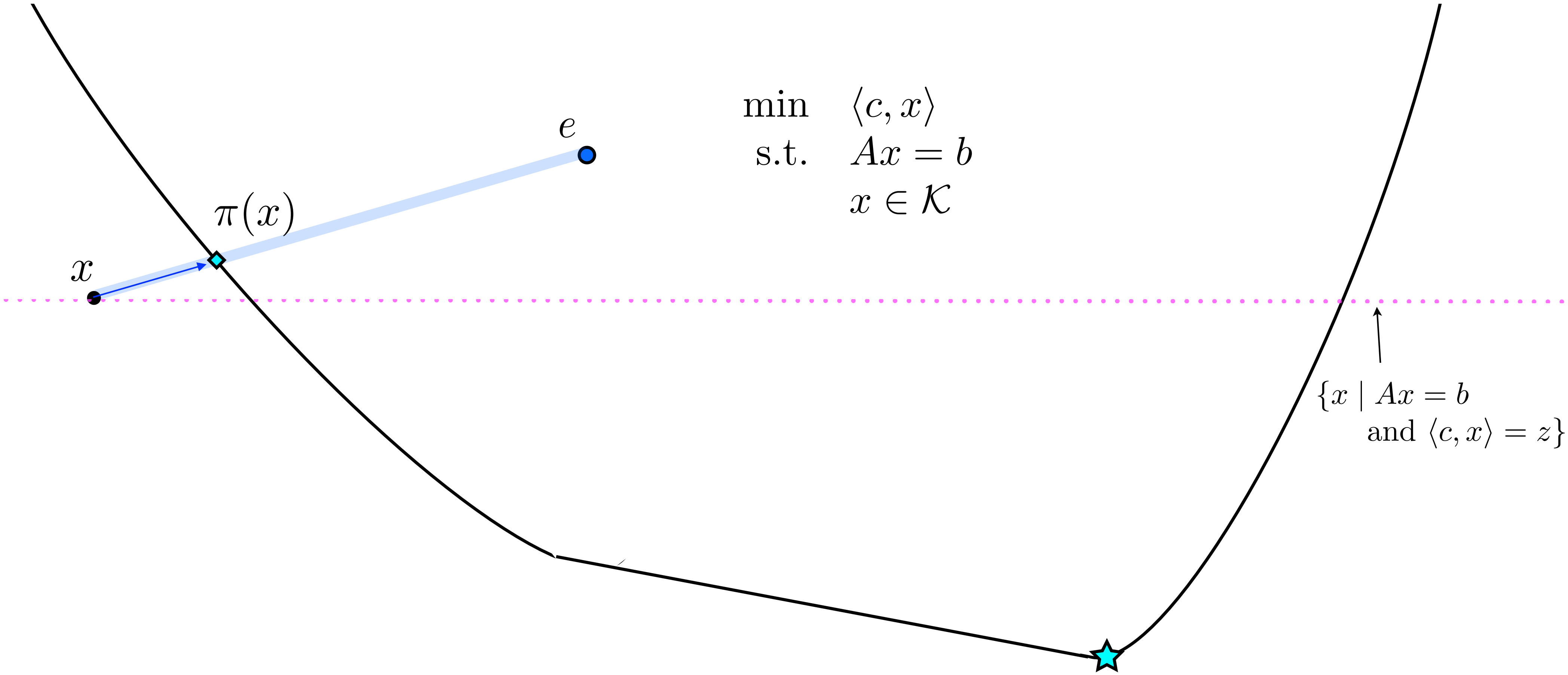
$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K} \end{aligned}$$

Assume the objective function $x \mapsto \langle c, x \rangle$ is constant on horizontal slices.

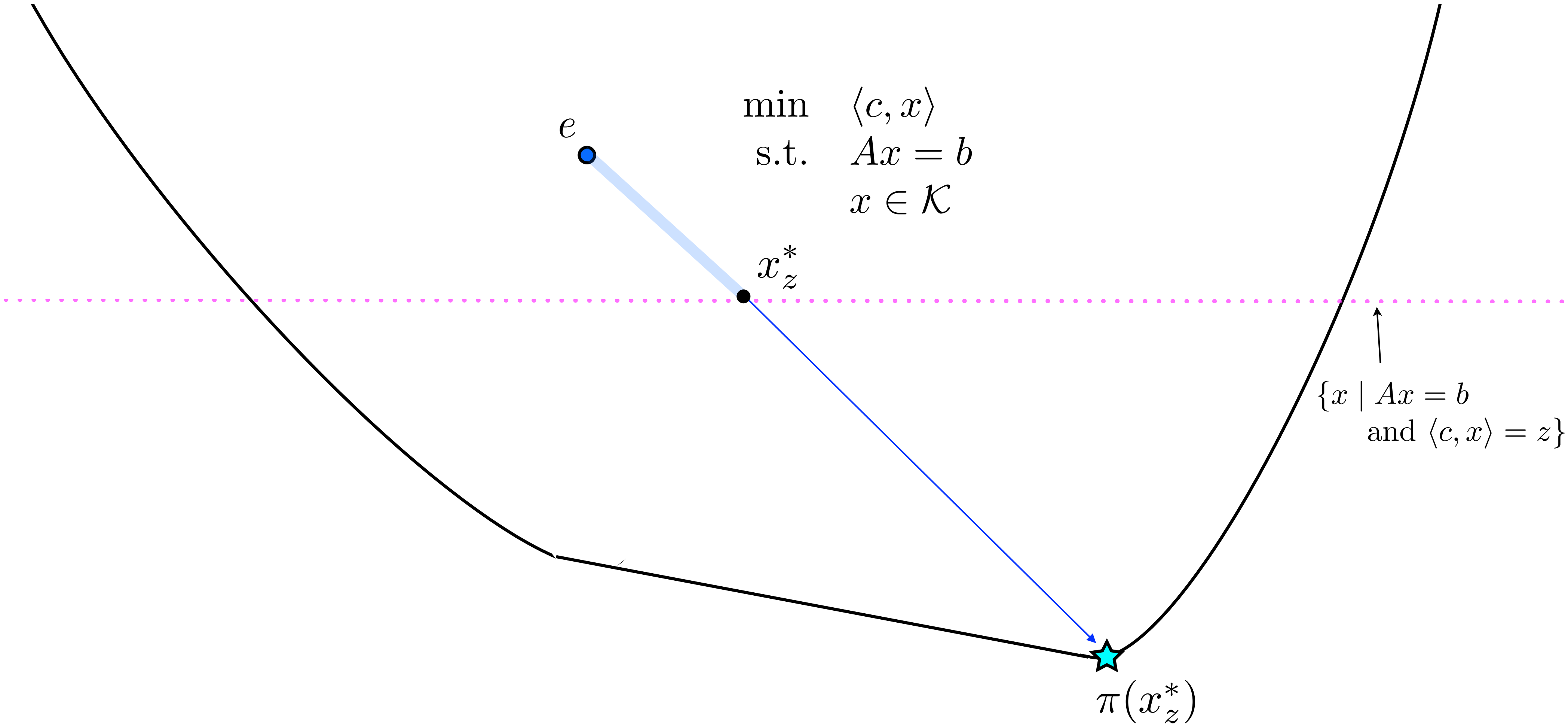

e
$\min \ \langle c, x \rangle$
$\text{s.t.} \ Ax = b$
$x \in \mathcal{K}$
$\{x \mid Ax = b$
$\text{and } \langle c, x \rangle = z\}$



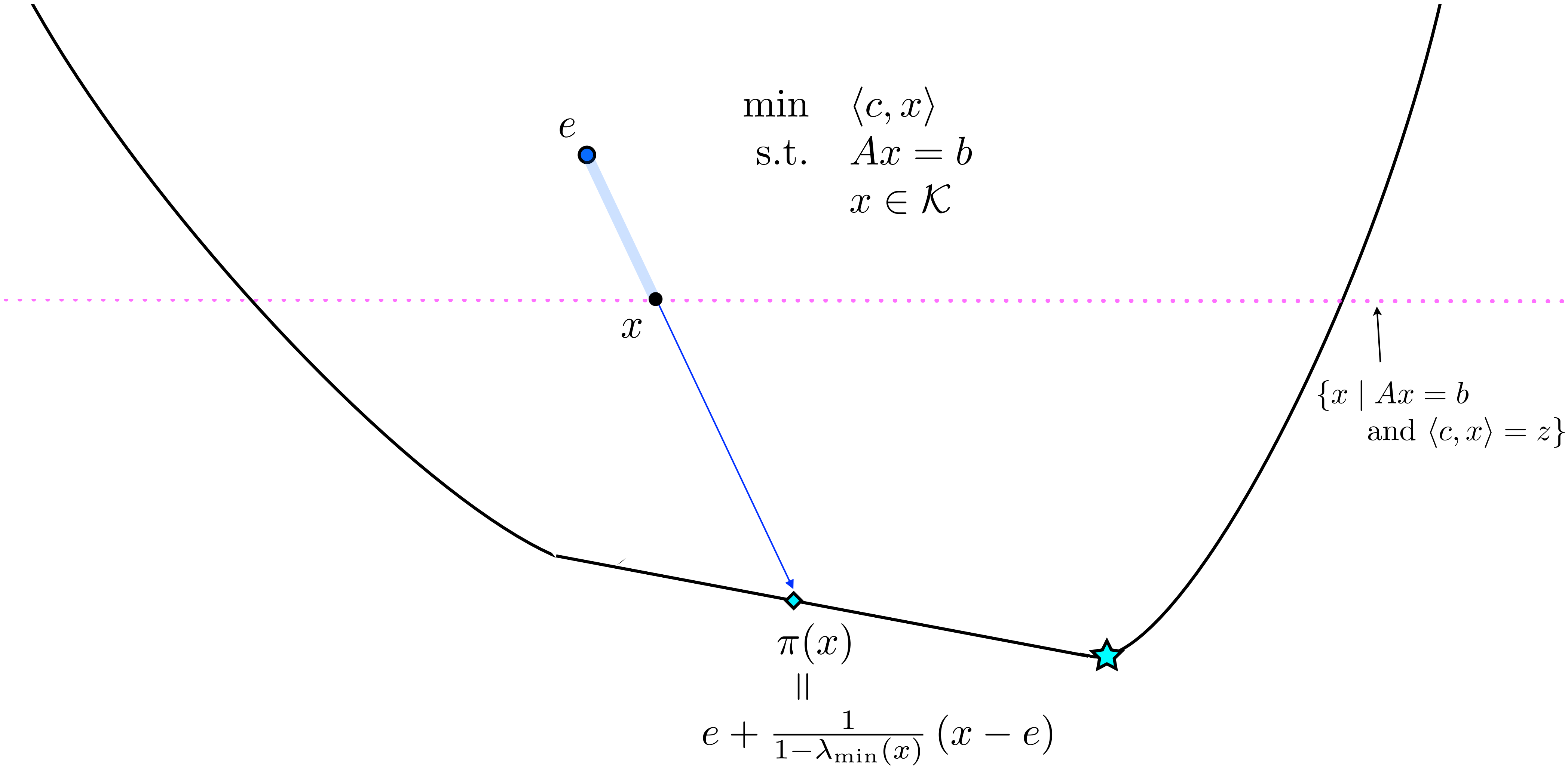
min $\langle c, x \rangle$
s.t. $Ax = b$
$x \in \mathcal{K}$
e
x
$\pi(x)$
$\{x \mid Ax = b$
and $\langle c, x \rangle = z\}$

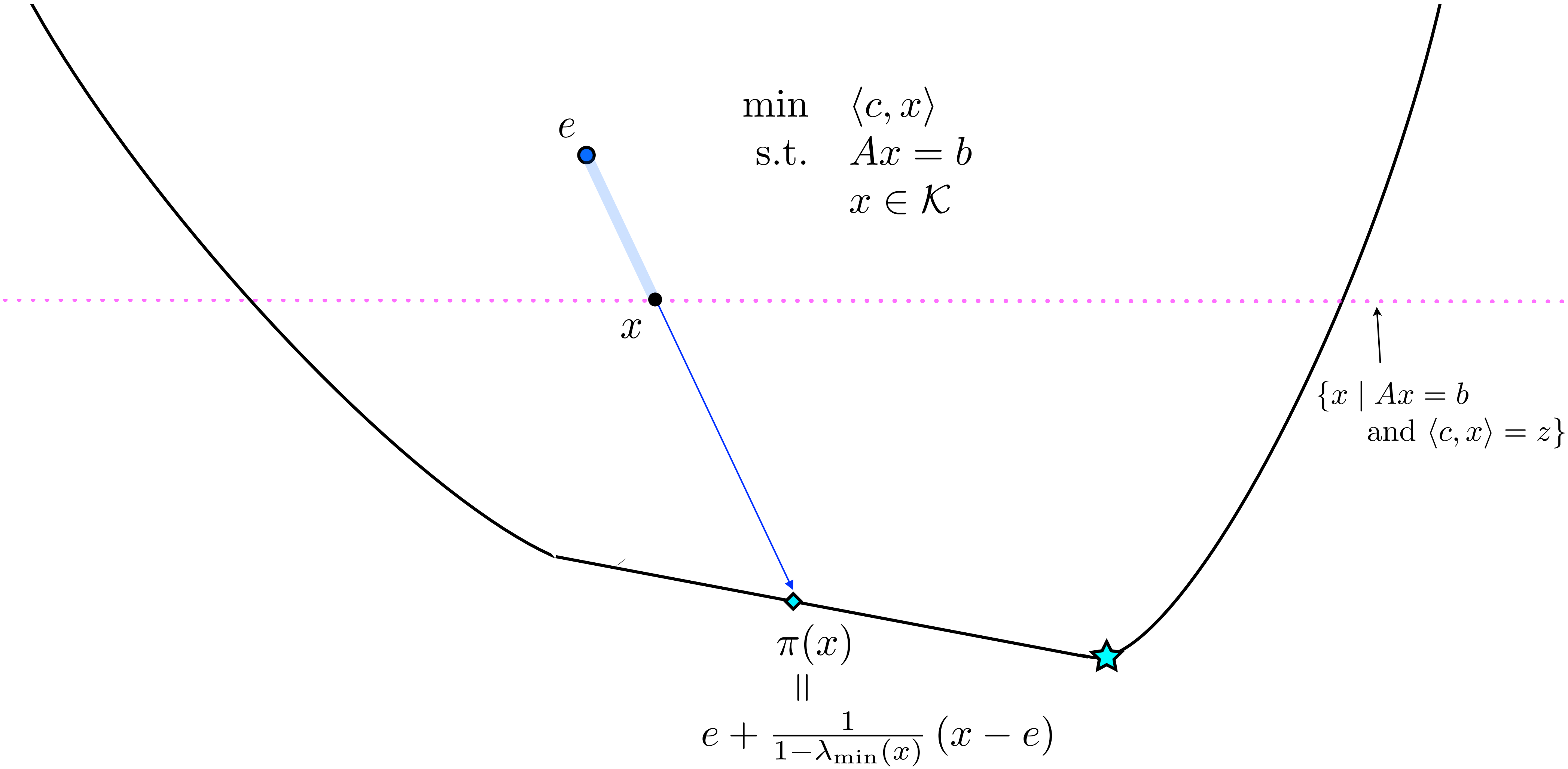


e
$\pi(x)$
x
$$\begin{array}{rl} \min & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}$$
$\{x \mid Ax = b$ and $\langle c, x \rangle = z\}$

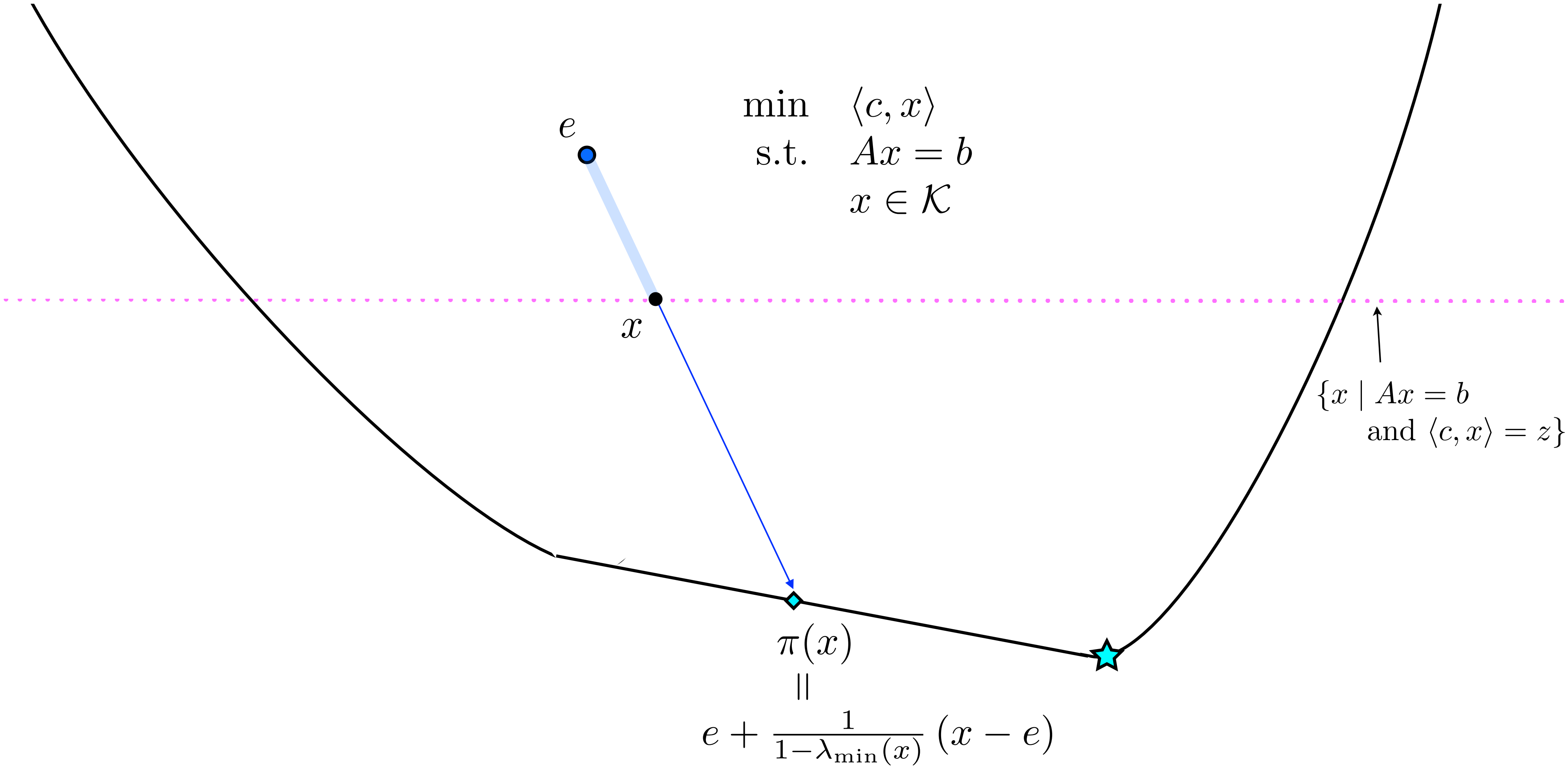


$$\begin{array}{rl} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}$$
e
x_z^*
$\{x \mid Ax = b$ and $\langle c, x\rangle = z\}$
$\pi(x_z^*)$


$\min \quad \langle c, x\rangle$
$\text{s.t.} \quad Ax = b$
$x \in \mathcal{K}$
e
x
$\{x \mid Ax = b$
$\text{and } \langle c, x\rangle = z\}$
$\pi(x)$
$||$
$e + \frac{1}{1-\lambda_{\min}(x)}(x-e)$



$$\langle c, \pi(x)\rangle = \langle c, e\rangle + \frac{1}{1-\lambda_{\min}(x)}(\langle c, x\rangle - \langle c, e\rangle)$$



$$\begin{aligned}\langle c, \pi(x)\rangle &= \langle c, e\rangle + \tfrac{1}{1-\lambda_{\min}(x)}(\langle c, x\rangle - \langle c, e\rangle) \\ &= \langle c, e\rangle + \tfrac{1}{1-\lambda_{\min}(x)}\,\underbrace{(z - \langle c, e\rangle)}_{\text{a negative constant}}\end{aligned}$$

$$
\begin{array}{rl} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}
\quad \equiv \quad
\begin{array}{rl} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x\rangle = z \end{array}
$$

Applying the subgradient method
– rather, supgradient method –
results in a sequence $x_0, x_1, \ldots$
for which

$$\ell \geq \big(\mathrm{dist}(x_0, X_z^*)/(r\epsilon)\big)^2$$

$$\Rightarrow \max_{k \leq \ell} \lambda_{\min}(x_k) \geq \lambda_{\min}^* - \epsilon$$

$$\begin{array}{rl} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array} \quad \equiv \quad \begin{array}{rl} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x\rangle = z \end{array}$$

Applying the subgradient method – rather, supgradient method – results in a sequence $x_0, x_1, \ldots$ for which

$$\ell \geq \left(\operatorname{dist}(x_0, X_z^*)/(r\epsilon)\right)^2$$
$$\Rightarrow \max_{k \leq \ell} \lambda_{\min}(x_k) \geq \lambda_{\min}^* - \epsilon$$

But what we would like is x_k for which $\langle c, \pi(x_k)\rangle \leq z^* + \epsilon$, where z^* is the optimal value of the hyperbolic program.

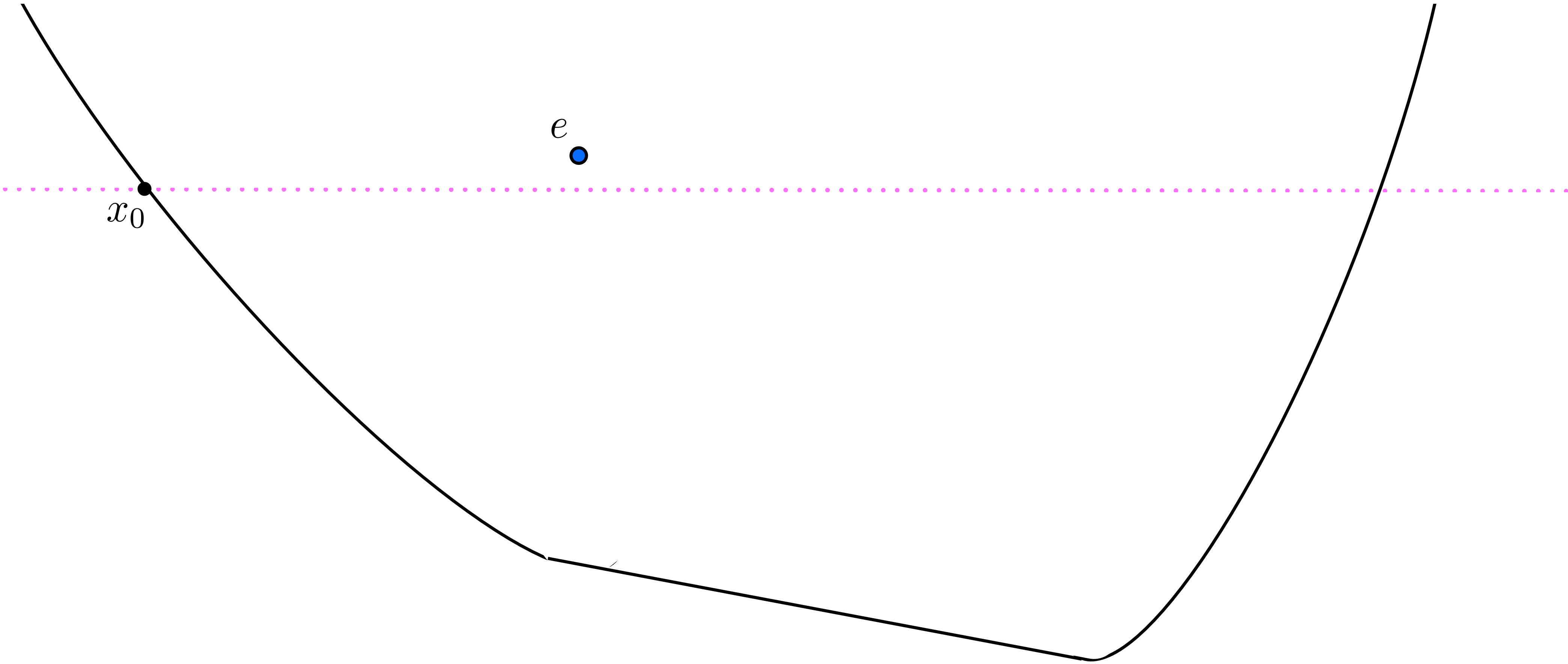
$$\begin{array}{ll} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array} \qquad \equiv \qquad \begin{array}{ll} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x\rangle = z \end{array}$$

Applying the subgradient method – rather, supgradient method – results in a sequence $x_0, x_1, \ldots$ for which

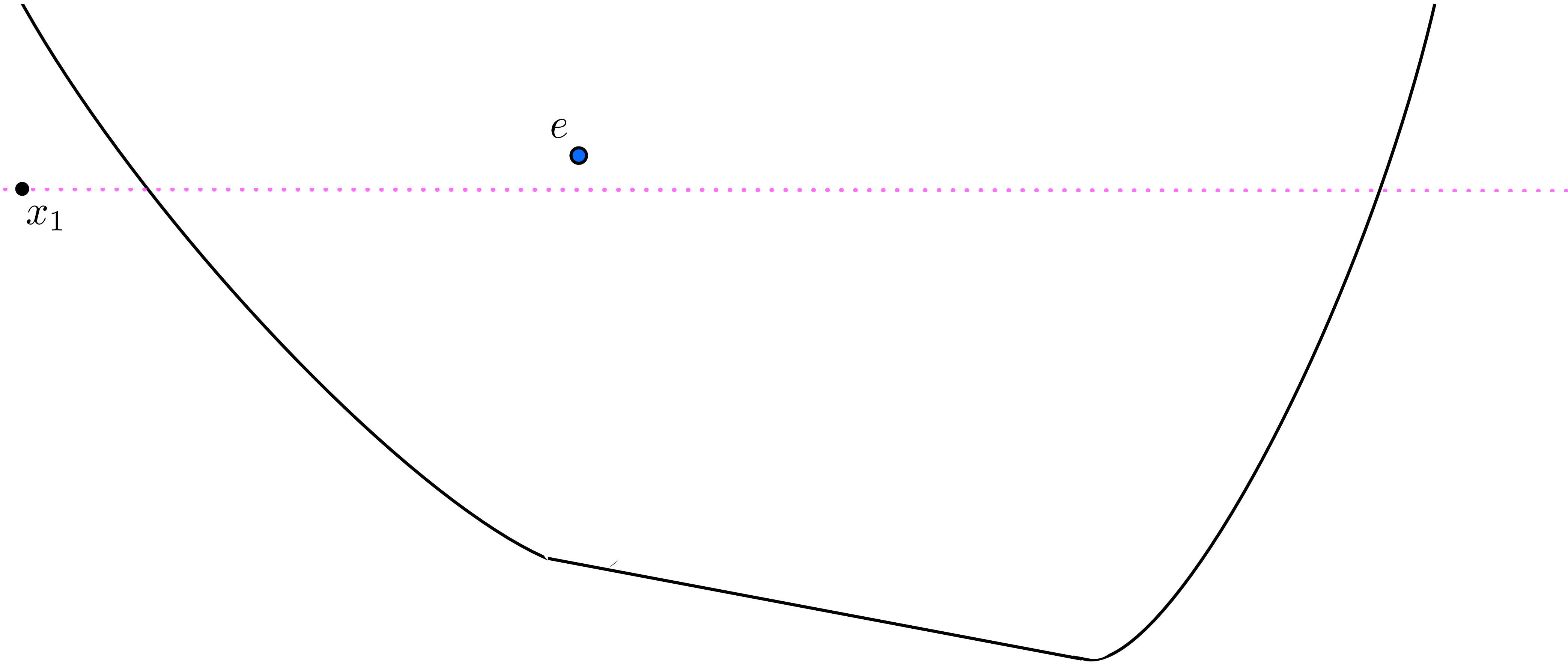
$$\ell \geq \big(\mathrm{dist}(x_0, X_z^*)/(r\epsilon)\big)^2$$
$$\Rightarrow \max_{k \leq \ell} \lambda_{\min}(x_k) \geq \lambda_{\min}^* - \epsilon$$

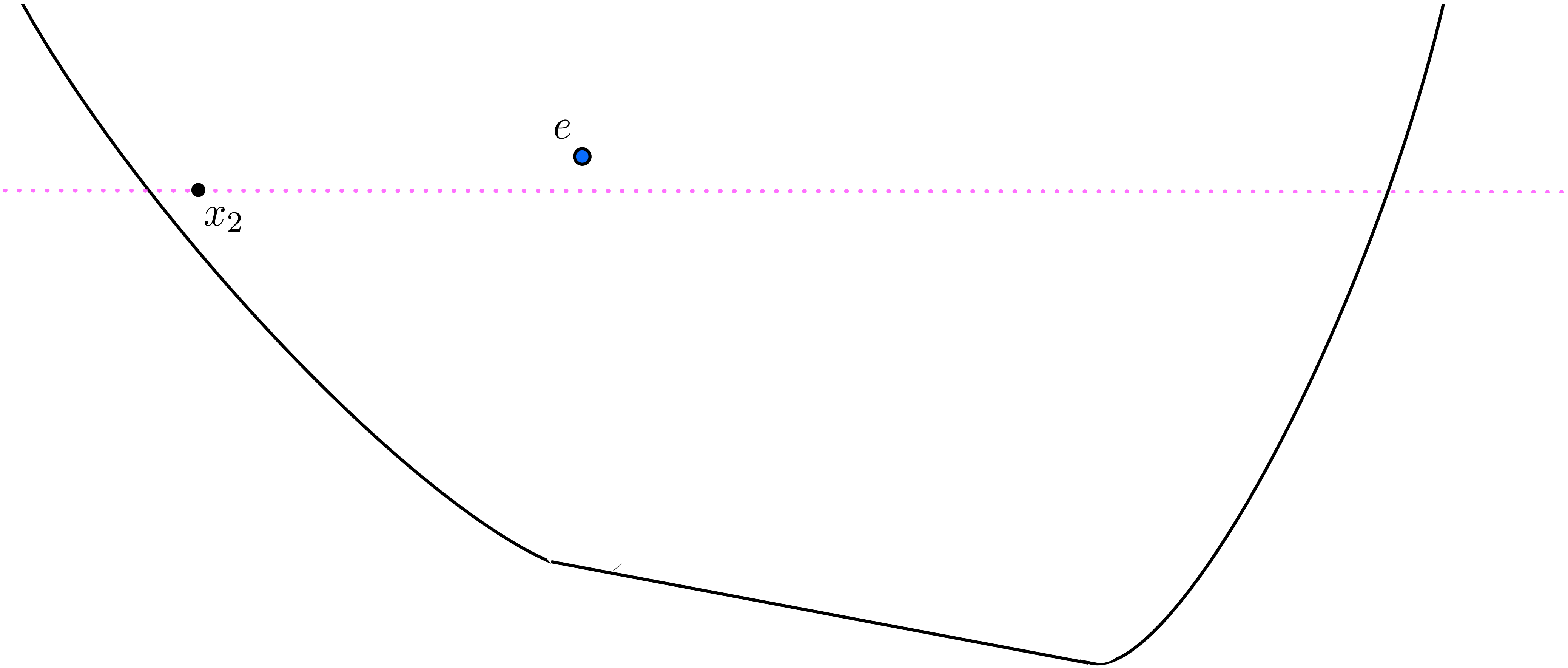
But what we would like is x_k for which $\langle c, \pi(x_k)\rangle \leq z^* + \epsilon$, where z^* is the optimal value of the hyperbolic program.

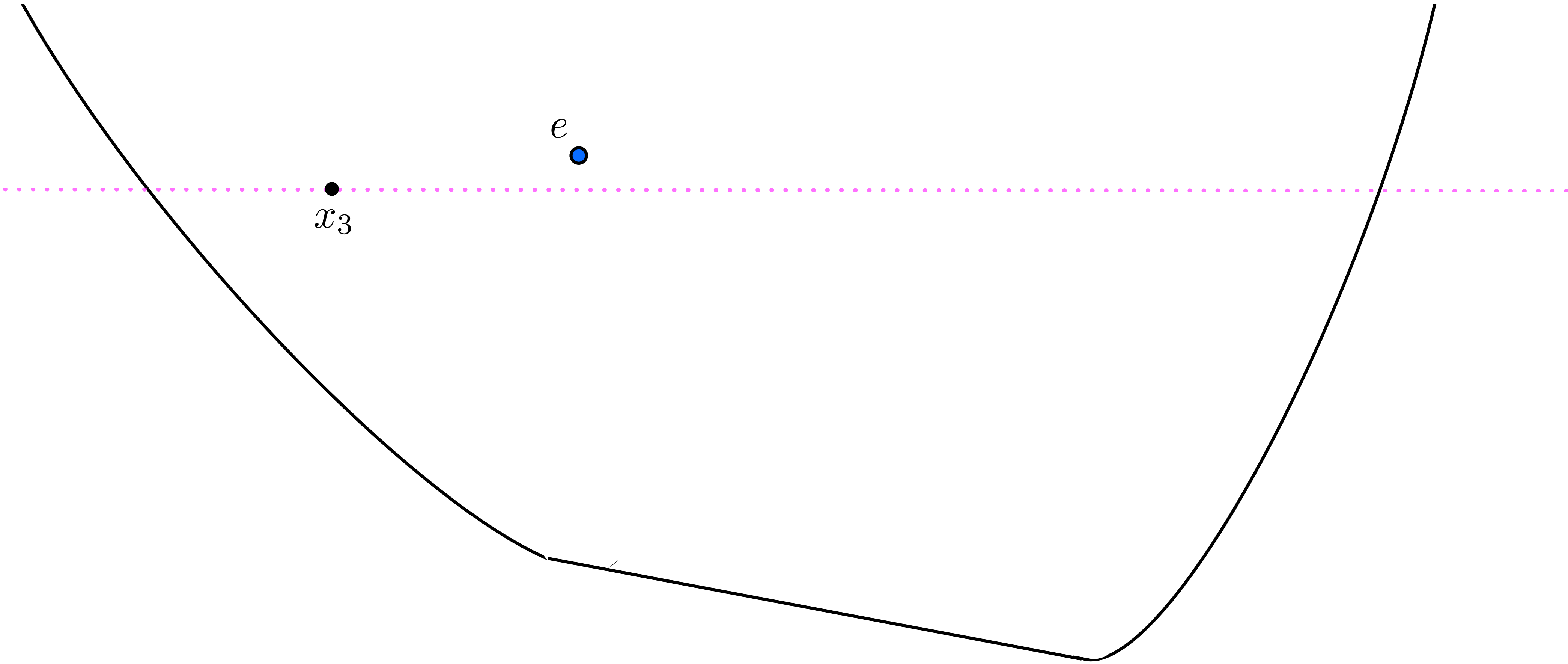
By relying on an approach which makes use of a sequence of values $z^{(j)}$ for the optimization problem on the right, we are able to devise an algorithm with the desired property, except for error being measured relatively rather than absolutely:

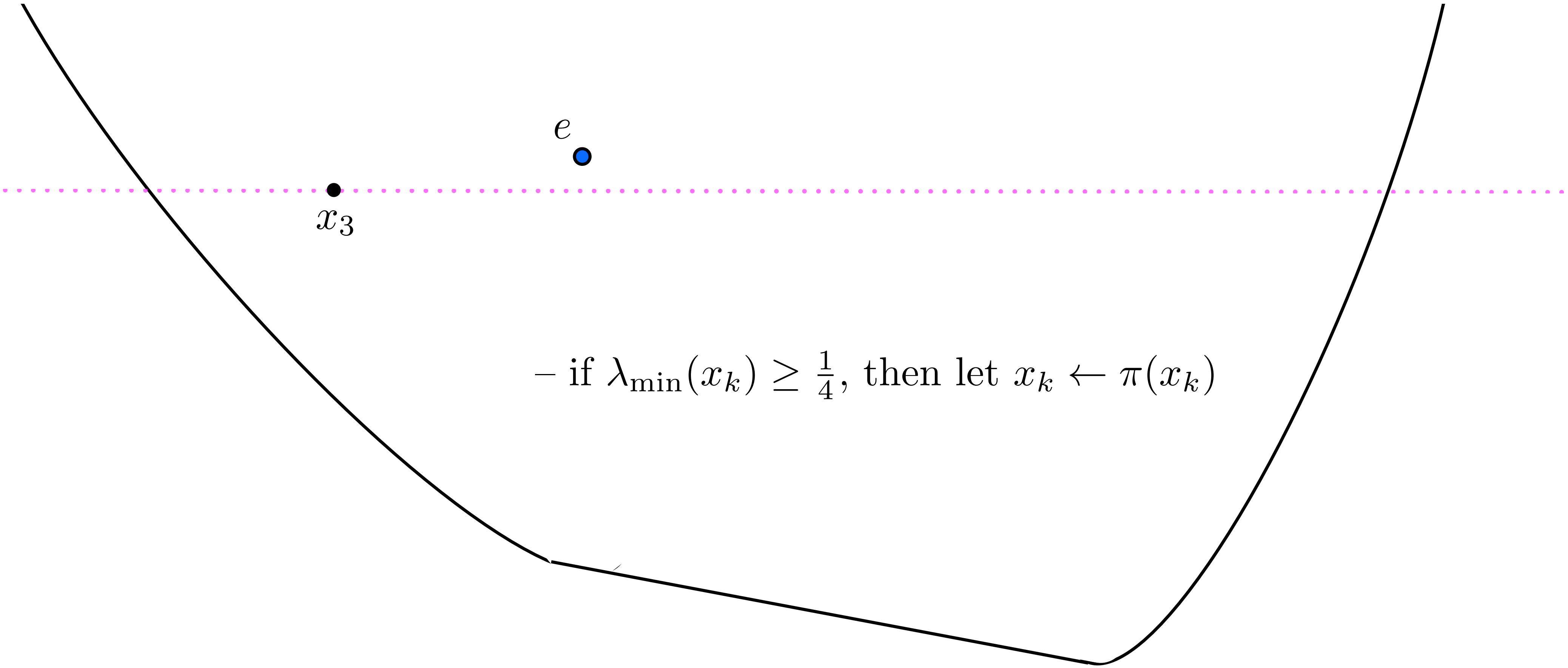


e
x_0

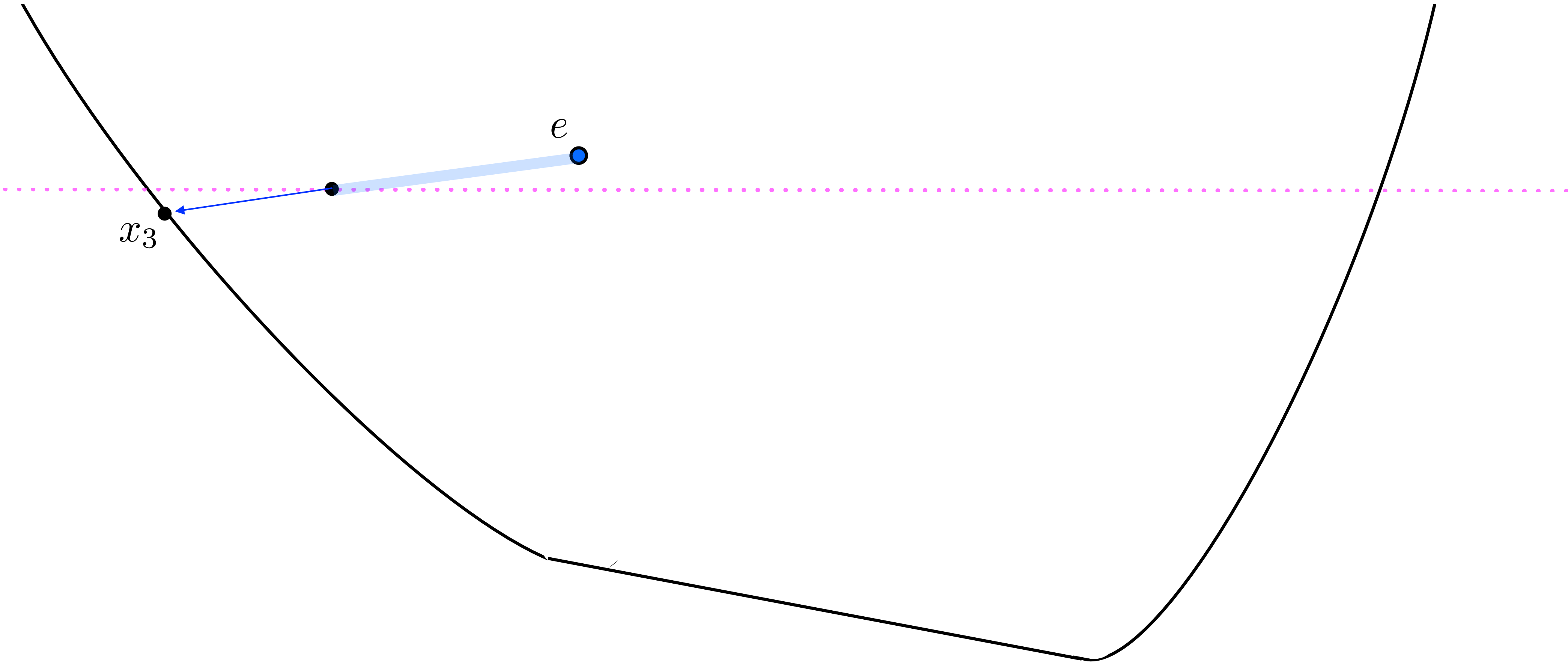

e
x_1


e
x_2

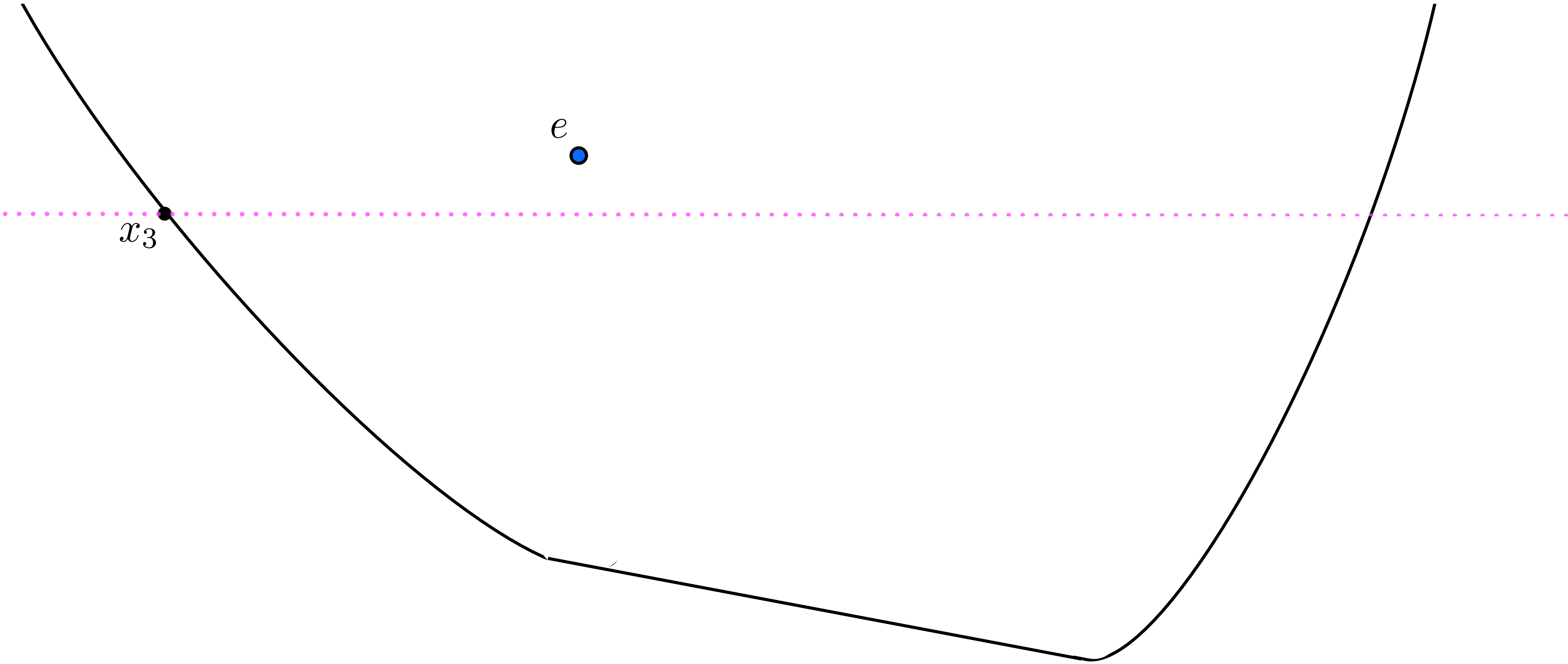

e
x_3

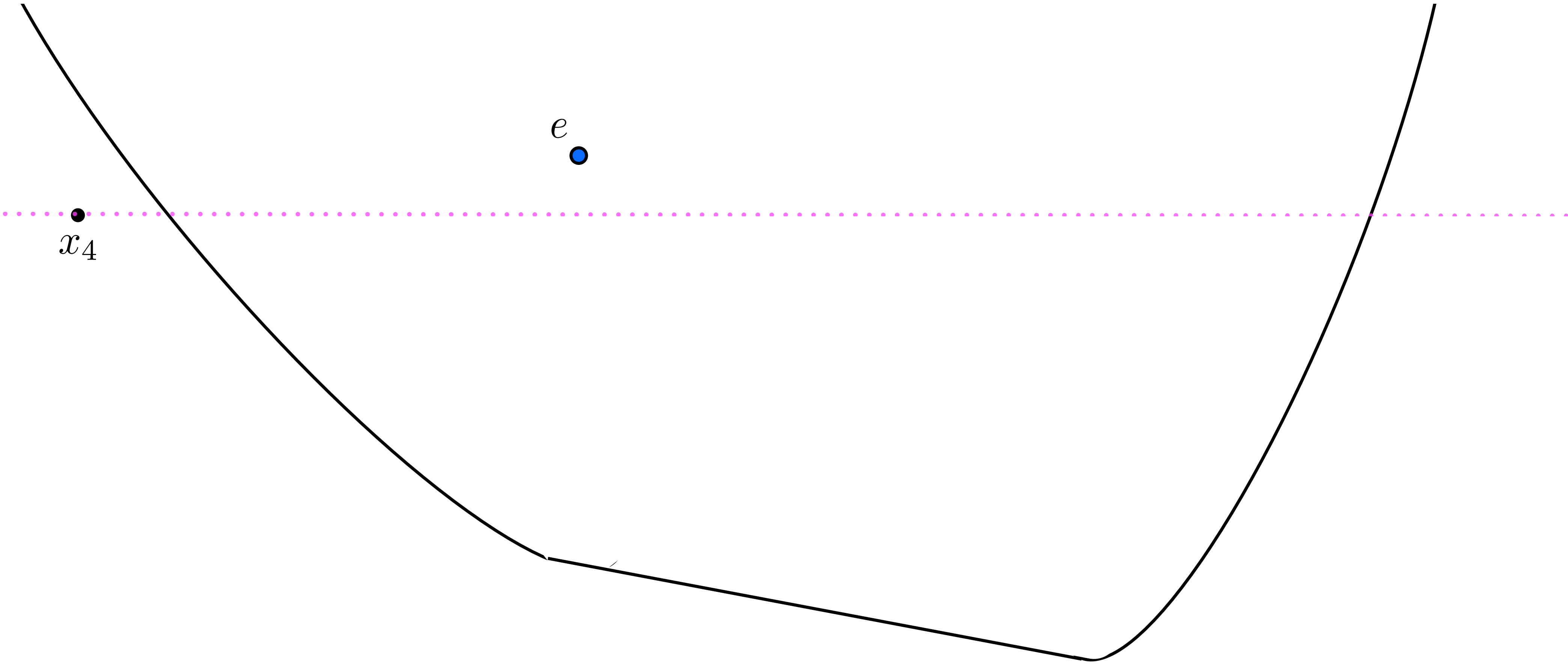


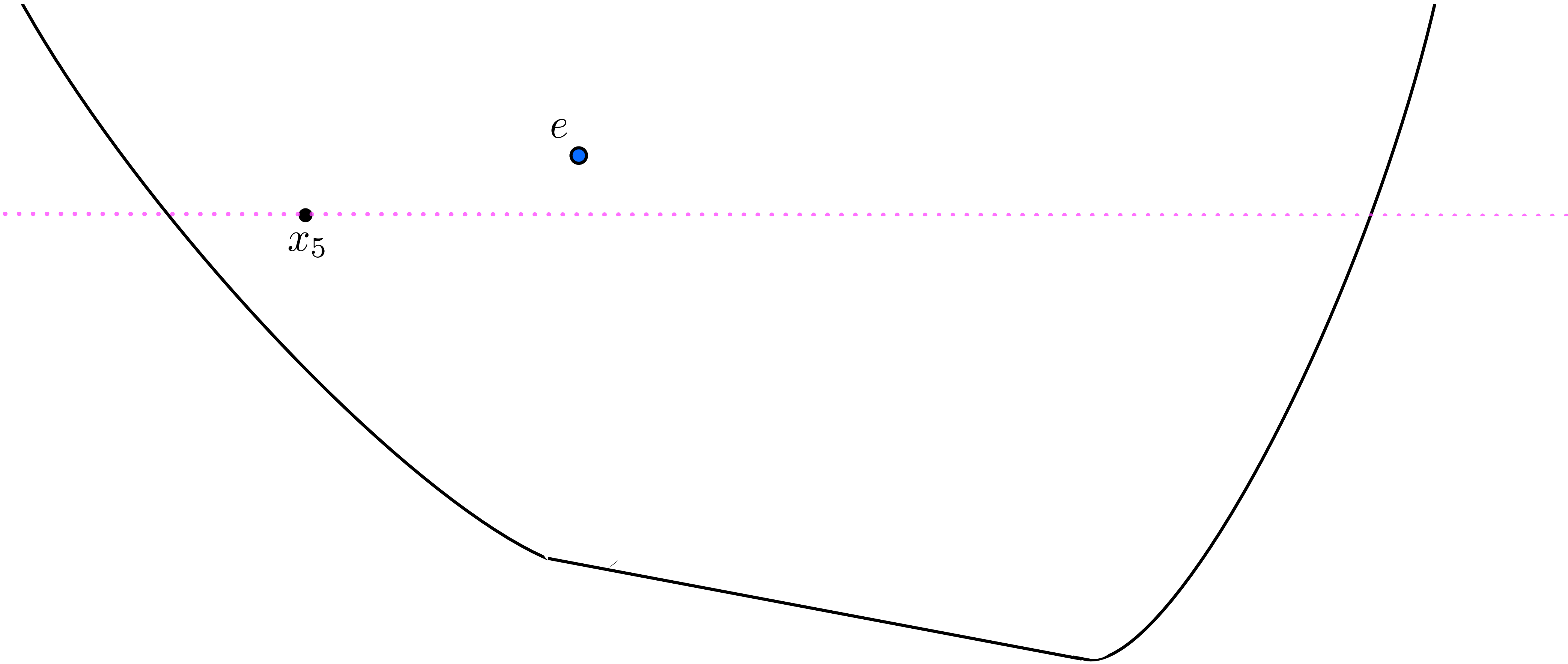
– if $\lambda_{\min}(x_k) \geq \frac{1}{4}$, then let $x_k \leftarrow \pi(x_k)$



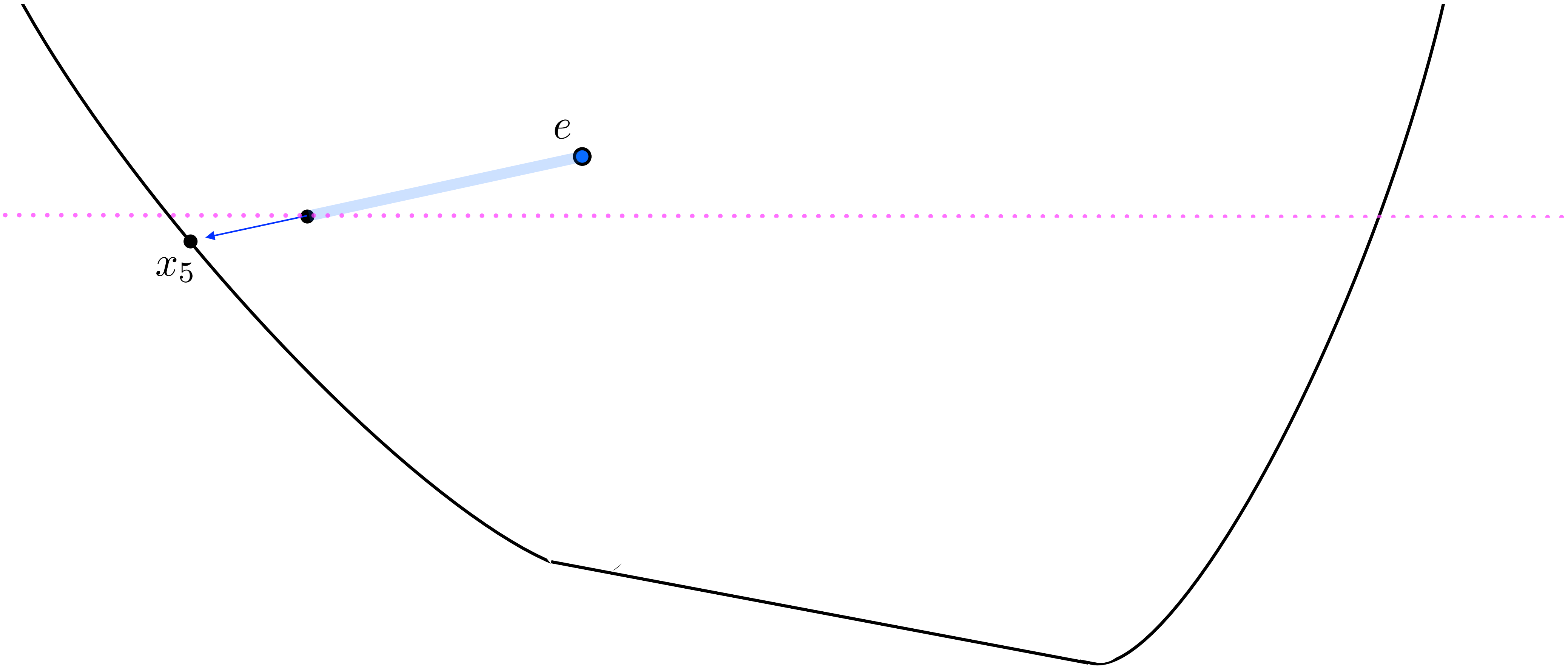
e
x_3


e
x_3

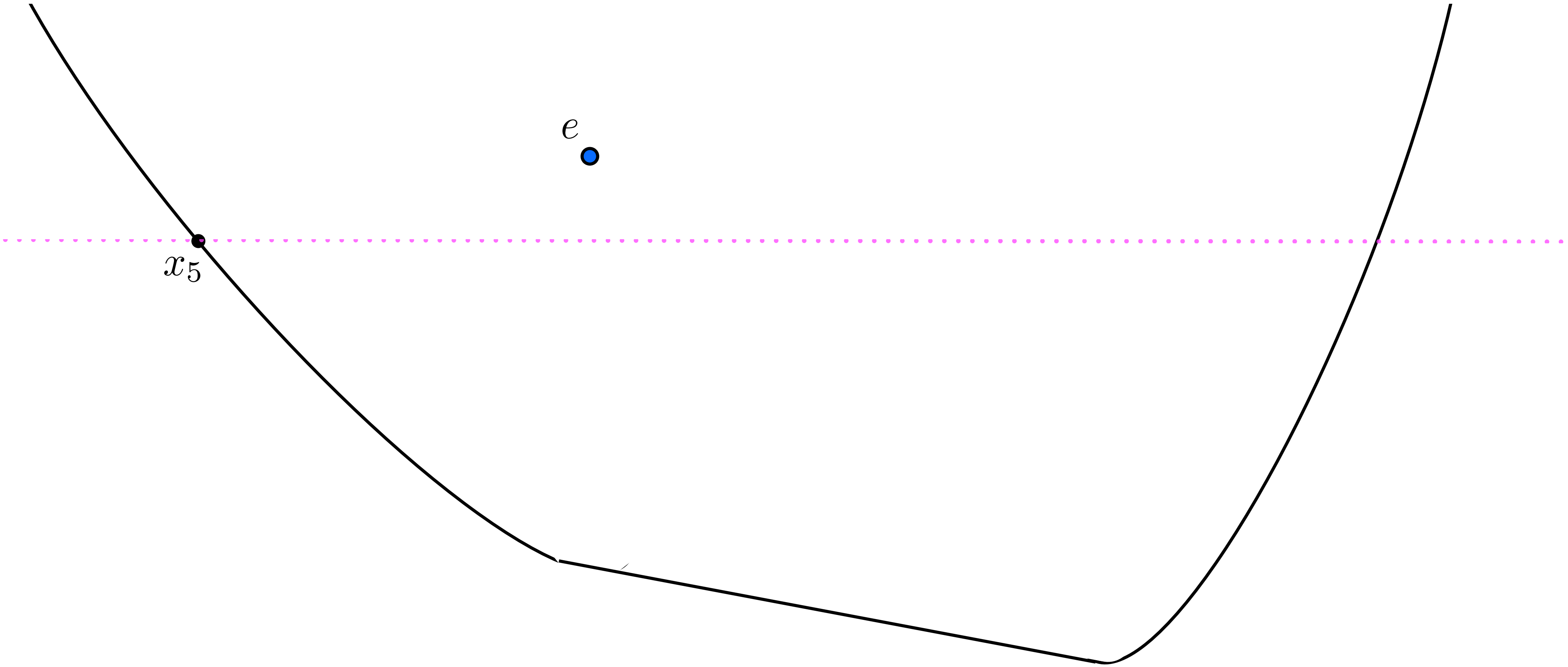

e
x4



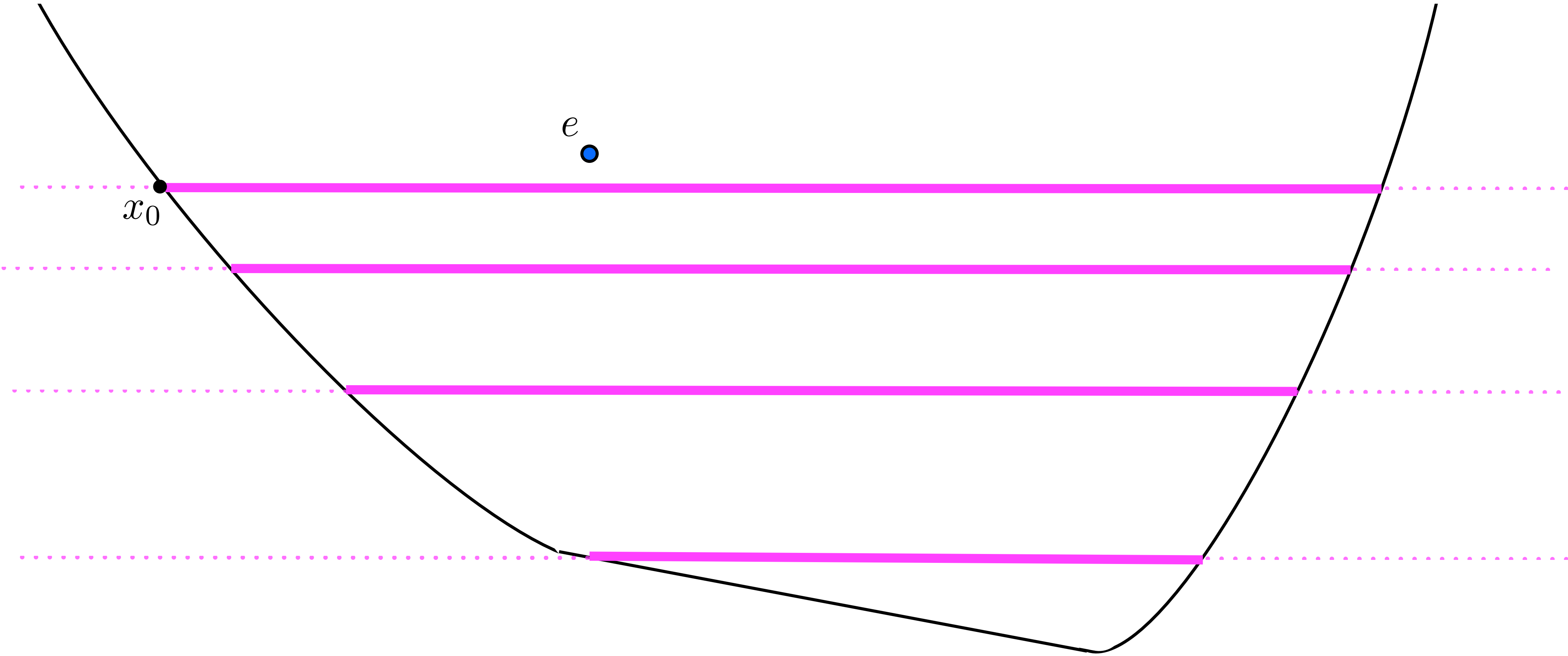
e
x_5



e
x_5



e
x_5



$=$ level set

$\mathrm{Diam} :=$ supremum of diameters of level sets for objective values $\leq c \cdot x_0$

$$\begin{array}{ll} \min & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array} \quad \equiv \quad \begin{array}{ll} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x \rangle = z \end{array}$$

Applying the subgradient method – rather, supgradient method – results in a sequence $x_0, x_1, \ldots$ for which

$$\ell \geq \big(\mathrm{dist}(x_0, X_z^*)/(r\epsilon)\big)^2$$
$$\Rightarrow \max_{k \leq \ell} \lambda_{\min}(x_k) \geq \lambda_{\min}^* - \epsilon$$

But what we would like is x_k for which $\langle c, \pi(x_k) \rangle \leq z^* + \epsilon$, where z^* is the optimal value of the hyperbolic program.

By relying on an approach which makes use of a sequence of values $z^{(j)}$ for the optimization problem on the right, we are able to devise an algorithm with the desired property, except for error being measured relatively rather than absolutely:

$$\ell \geq 8 \left(\frac{\mathrm{Diam}}{r_e} \right)^2 \cdot \left(\frac{1}{\epsilon^2} + \frac{1}{\epsilon} \log_{4/3} \left(\frac{\langle c, e \rangle - z^*}{\langle c, e \rangle - z^{(0)}} \right) + 1 \right)$$

$$\Rightarrow \quad \min_{k \leq \ell} \frac{\langle c, \pi(x_k) \rangle - z^*}{\langle c, e \rangle - z^*} \leq \epsilon$$

error measurement is relative rather than absolute

Subsequently, PhD student Ben Grimmer designed a more attractive algorithm and obtained a bound with the same dependence on ϵ, but nicer in other regards.

Smoothing

Motivated by work of Nesterov pertaining to SDP,
we rely on the concave function

$$\boxed{f_\mu(x) := -\mu \, \ln \textstyle\sum_j \exp(-\lambda_j(x)/\mu) \quad \text{(for fixed } \mu > 0)}$$

Easy to see: $\lambda_{\min}(x) - \mu \ln n \leq f_\mu(x) \leq \lambda_{\min}(x)$

– thus, if $\mu = \epsilon/(2 \ln n)$ then $\lambda_{\min}(x) - \frac{\epsilon}{2} \leq f_\mu(x) \leq \lambda_{\min}(x)$

Prop: f is analytic and $\|\nabla f_\mu(x) - \nabla f_\mu(y)\|_\infty^* \leq \frac{1}{\mu}\|x - y\|_\infty$

Pf: Thanks to Nesterov, Helton, Vinnikov and an old analysis result. $\square$

Cor: $\|\nabla f_\mu(x) - \nabla f_\mu(y)\| \leq \frac{1}{r_e^2 \mu}\|x - y\|$ (Euclidean norm)

In some important cases (e.g., $\mathbb{R}^n_+$) the value r_e is easily computed,
but it not realistic to assume r_e is easily computable
when $\mathcal{K}$ is a general hyperbolicity cone.

Thus we rely on a"universal" accelerated method by Nesterov
which requires only a guess of the Lipschitz constant ...

Thm: Obtain iterate x_k satisfying $\dfrac{\langle c, \pi(x_k)\rangle - z^*}{\langle c, e\rangle - z^*} \leq \epsilon$

with the number of gradient evaluations being only of order

$$\frac{\mathrm{Diam}}{\epsilon}\sqrt{\frac{\ln n}{r_e}} \qquad + \left(1 + \log_2 \frac{\langle c, e\rangle - z^*}{\langle c, e\rangle - z^{(0)}}\right)\left(1 + \mathrm{Diam}\sqrt{\frac{\ln n}{r_e}} + \left|\log_2 \frac{L^\bullet}{L}\right|\right)$$

The paper also gives focus to showing that if values and gradients can be efficiently computed for p, then they can be computed efficiently for f_μ

This is made explicit for cones which are intersections of quadratic cones.

Two facts helpful in motivating what follows:

1) For SDP,
computing $\nabla f_\mu(X)$ requires a full eigen-decomposition of the matrix X,
whereas for computing a supgradient of the function $X \mapsto \lambda_{\min}(X)$,
it suffices to compute (to high precision) an eigenvector only for $\lambda_{\min}(X)$.

$$\nabla f_\mu(X) = \frac{1}{\sum_{j=1}^n \exp(-\lambda_j(X)/\mu)} \; Q \begin{bmatrix} \exp(-\lambda_1(X)/\mu) & & \\ & \ddots & \\ & & \exp(-\lambda_n(X)/\mu) \end{bmatrix} Q^T$$

$$\text{where} \quad Q \begin{bmatrix} \lambda_1(X) & & \\ & \ddots & \\ & & \lambda_n(X) \end{bmatrix} Q^T \text{ is an eigendecomposition of } X$$

Two facts helpful in motivating what follows:

1) For SDP,
computing $\nabla f_\mu(X)$ requires a full eigen-decomposition of the matrix X,
whereas for computing a supgradient of the function $X \mapsto \lambda_{\min}(X)$,
it suffices to compute (to high precision) an eigenvector only for $\lambda_{\min}(X)$.

2) For a generic set of SDP's, there is a unique optimal solution X^*,
and the objective function at feasible points in a neighborhood of X^*
grows quadratically in the distance to X^*:

$$X \text{ feasible and } \|X - X^*\| \leq \delta \quad \Rightarrow \quad \langle C, X \rangle \geq \mu \|X - X^*\|^2$$

This fact has been available for almost 20 years,
but recently PhD student Lijun Ding was the first
to provide characterizations of μ and δ in terms natural to the SDP literature.

Two facts helpful in motivating what follows:

1) For SDP,
computing $\nabla f_\mu(X)$ requires a full eigen-decomposition of the matrix X,
whereas for computing a supgradient of the function $X \mapsto \lambda_{\min}(X)$,
it suffices to compute (to high precision) an eigenvector only for $\lambda_{\min}(X)$.

2) For a generic set of SDP's, there is a unique optimal solution X^*,
and the objective function at feasible points in a neighborhood of X^*
grows quadratically in the distance to X^*:

$$X \text{ feasible and } \|X - X^*\| \leq \delta \quad \Rightarrow \quad \langle C, X\rangle \geq \mu \|X - X^*\|^2$$

This fact has been available for almost 20 years,
but recently PhD student Lijun Ding was the first
to provide characterizations of μ and δ in terms natural to the SDP literature.

$$\begin{array}{rl} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x\rangle = z \end{array}$$

From **(2)** follows for each of the generic SDP's
that there exist positive δ_z and μ_z for which

$$x \text{ feasible and } \|x - x_z^*\| \leq \delta_z \quad \Rightarrow \quad \lambda_{\min}(x_z^*) - \lambda_{\min}(x) \geq \mu_z \|x - x_z^*\|^2$$

Now return to the setting for which the talk began ...

convex function

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in Q \end{array}$$

closed convex set

"step size"

$$x_{k+1} = P_Q(x_k - \alpha_k g_k)$$

orthogonal projection onto Q

gradient (or subgradient) of f at x_k

For non-differentiable f, we relied on the subgradient method, in which $\alpha_k = \epsilon / \|g_k\|^2$

$$\ell \geq \big(M \,\mathrm{dist}(x_0, X^*)/\epsilon \big)^2 \quad \Rightarrow \quad \min_{k \leq \ell} f(x_k) \leq f^* + \epsilon$$

set of optimal solutions

optimal objective value

When ϵ is small and $f(x_0) \gg f^*$, the step size $\alpha_k = \epsilon/\|g_k\|^2$ makes slow progress, leading to $1/\epsilon^2$ in the complexity bound.

Idea:

Apply, in parallel, subgradient methods with stepsizes

$$\epsilon/\|g_k\|^2, \; 2\epsilon/\|g_k\|^2, \; 2^2\epsilon/\|g_k\|^2, \ldots, 2^N\epsilon/\|g_k\|^2 \quad \text{where } N \approx \log_2(1/\epsilon)$$

...

$\texttt{subgrad}_{\boxed{n}}$ $x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$

Let $\bar{\mathbf{x}}$ be a feasible point known to the user.

$\texttt{subgrad}_{\boxed{4}}$

$f(\bar{\mathbf{x}})$

$\texttt{subgrad}_{\boxed{3}}$

$f(\bar{\mathbf{x}})$

$\texttt{subgrad}_{\boxed{2}}$

$f(\bar{\mathbf{x}})$

$\texttt{subgrad}_{\boxed{1}}$

$f(\bar{\mathbf{x}})$

$\texttt{subgrad}_{\boxed{0}}$

$f(\bar{\mathbf{x}})$

$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$

$\texttt{subgrad}_{\boxed{4}}$

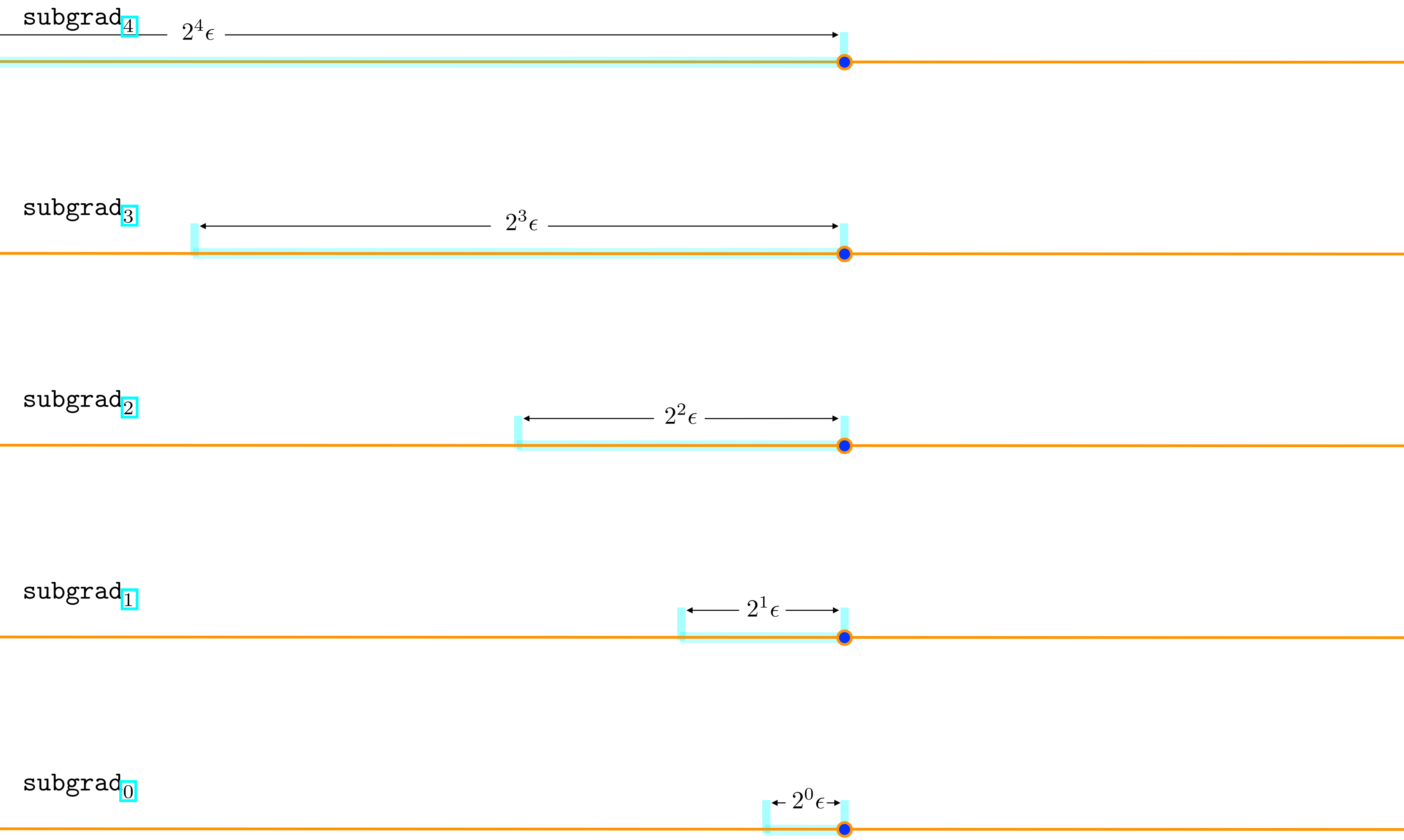
$\texttt{subgrad}_{\boxed{3}}$

$\texttt{subgrad}_{\boxed{2}}$

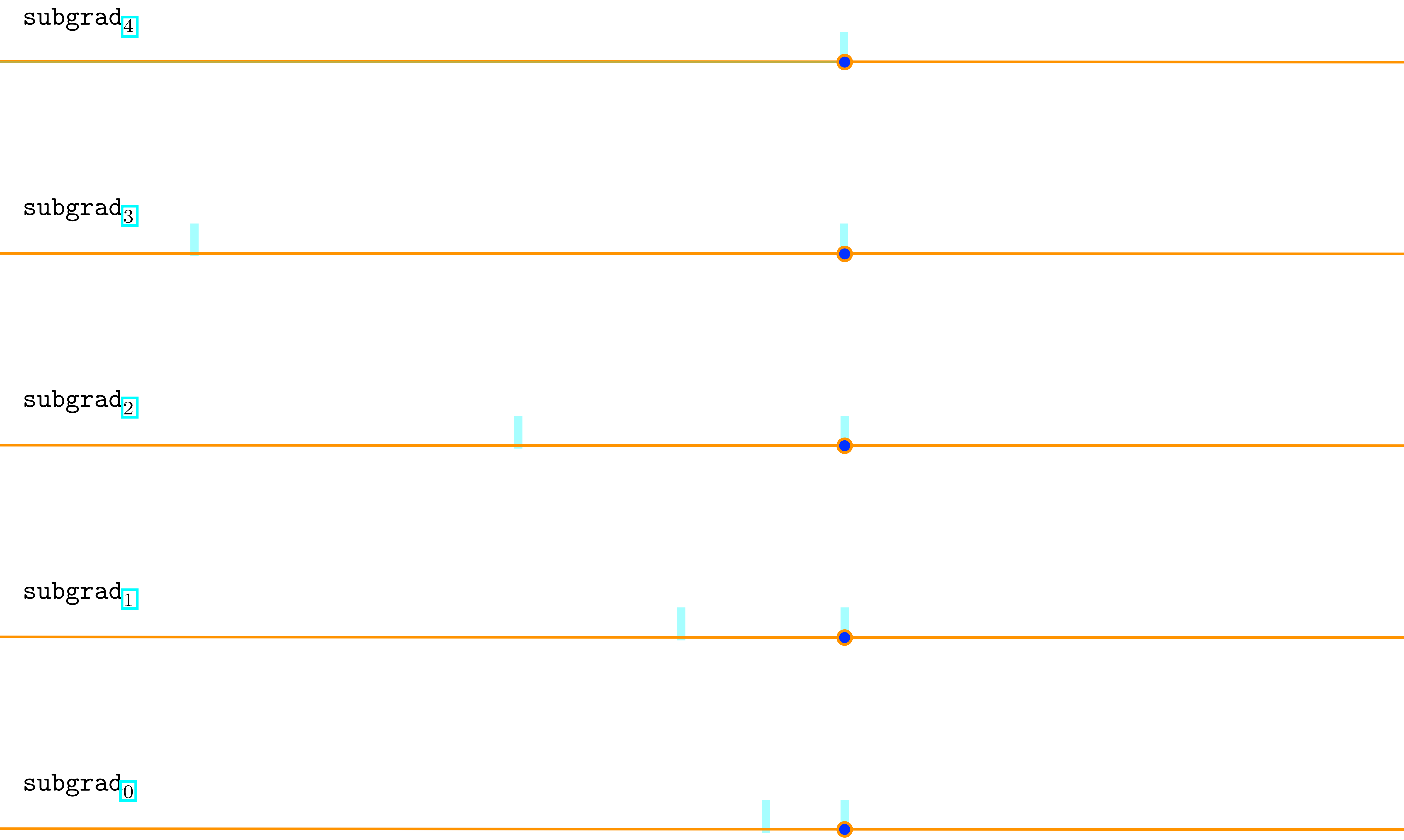
$\texttt{subgrad}_{\boxed{1}}$

$\texttt{subgrad}_{\boxed{0}}$

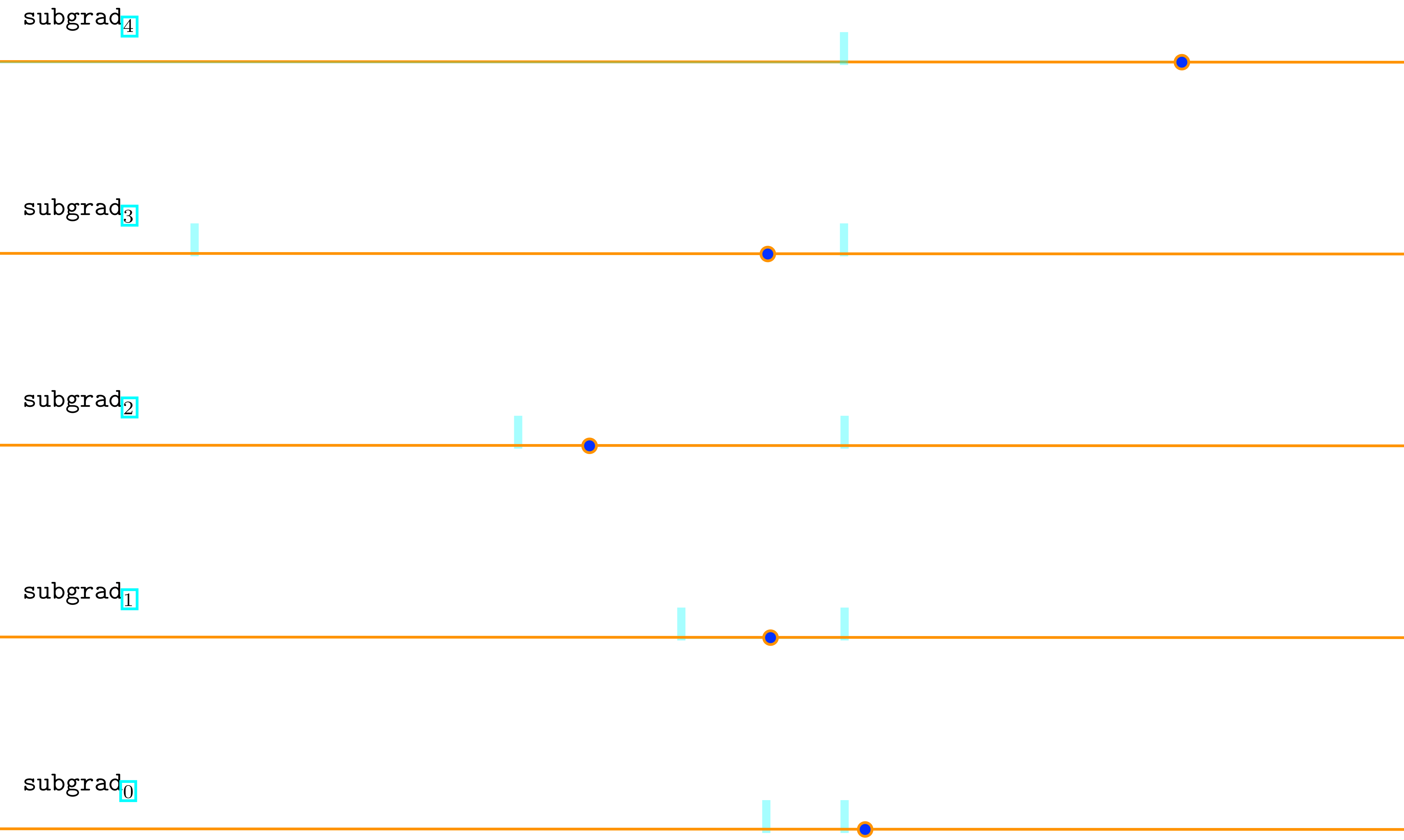
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n\epsilon}\, g/\|g\|^2\,)$



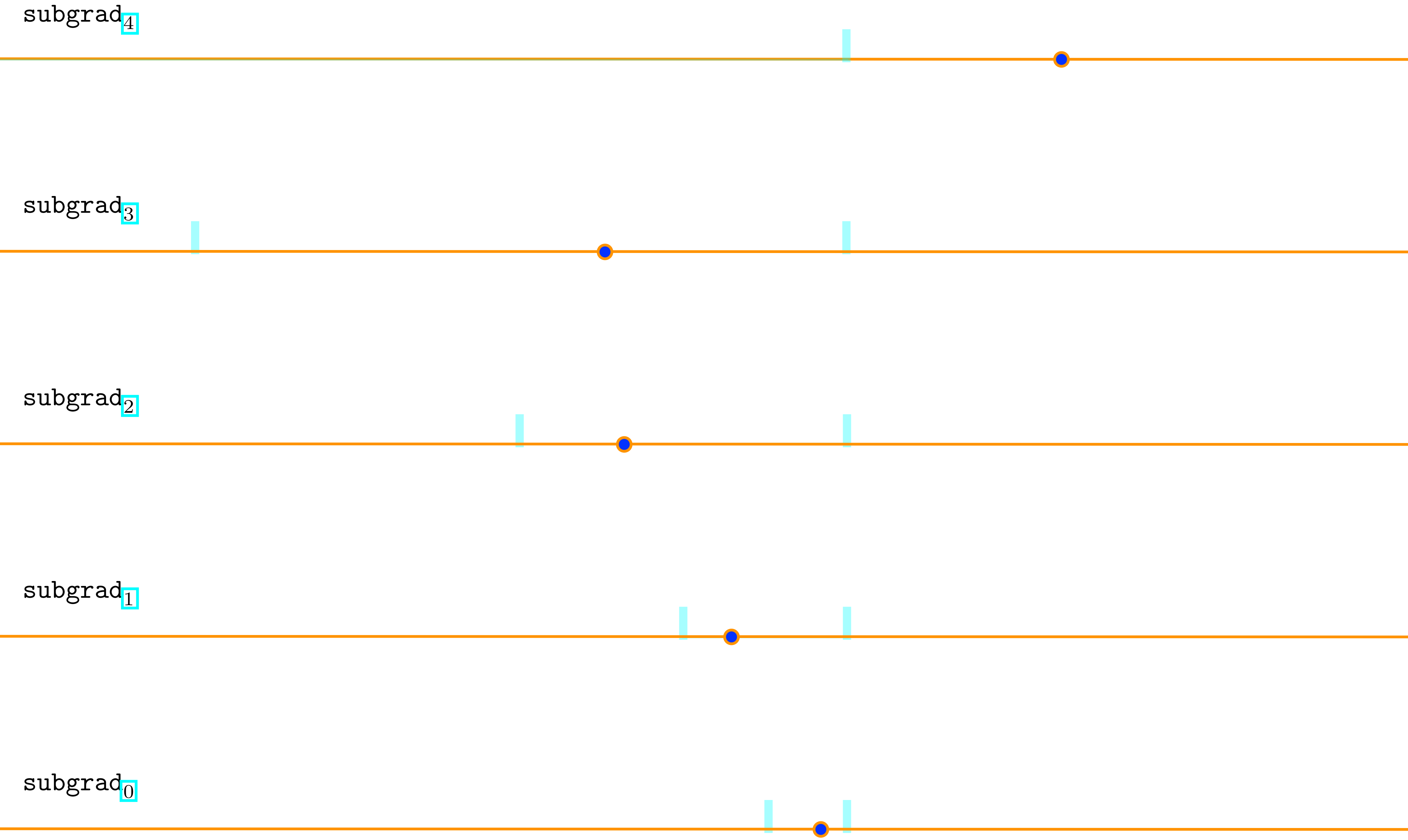
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\,x - \boxed{2^n\epsilon}\,g/\|g\|^2\,)$



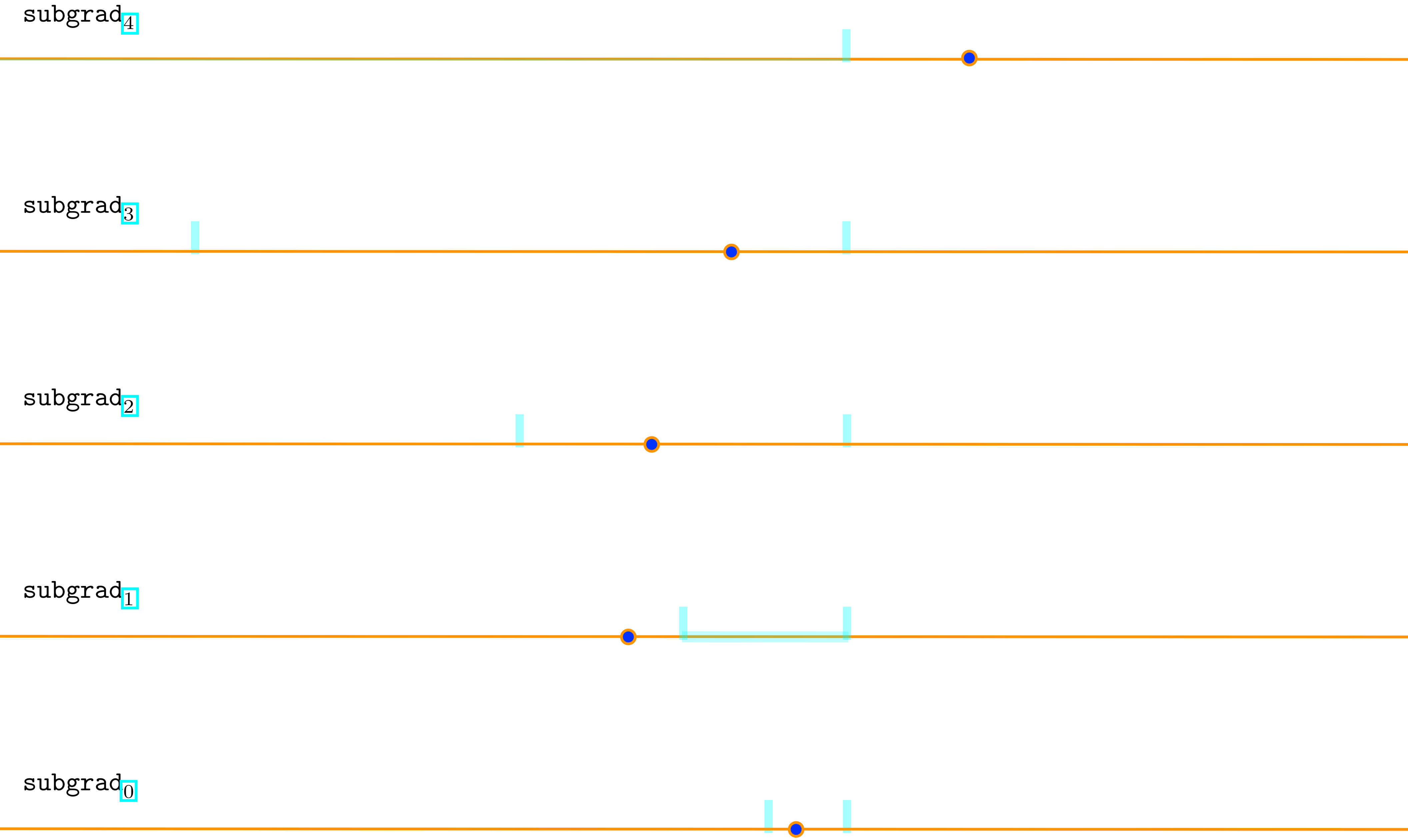
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



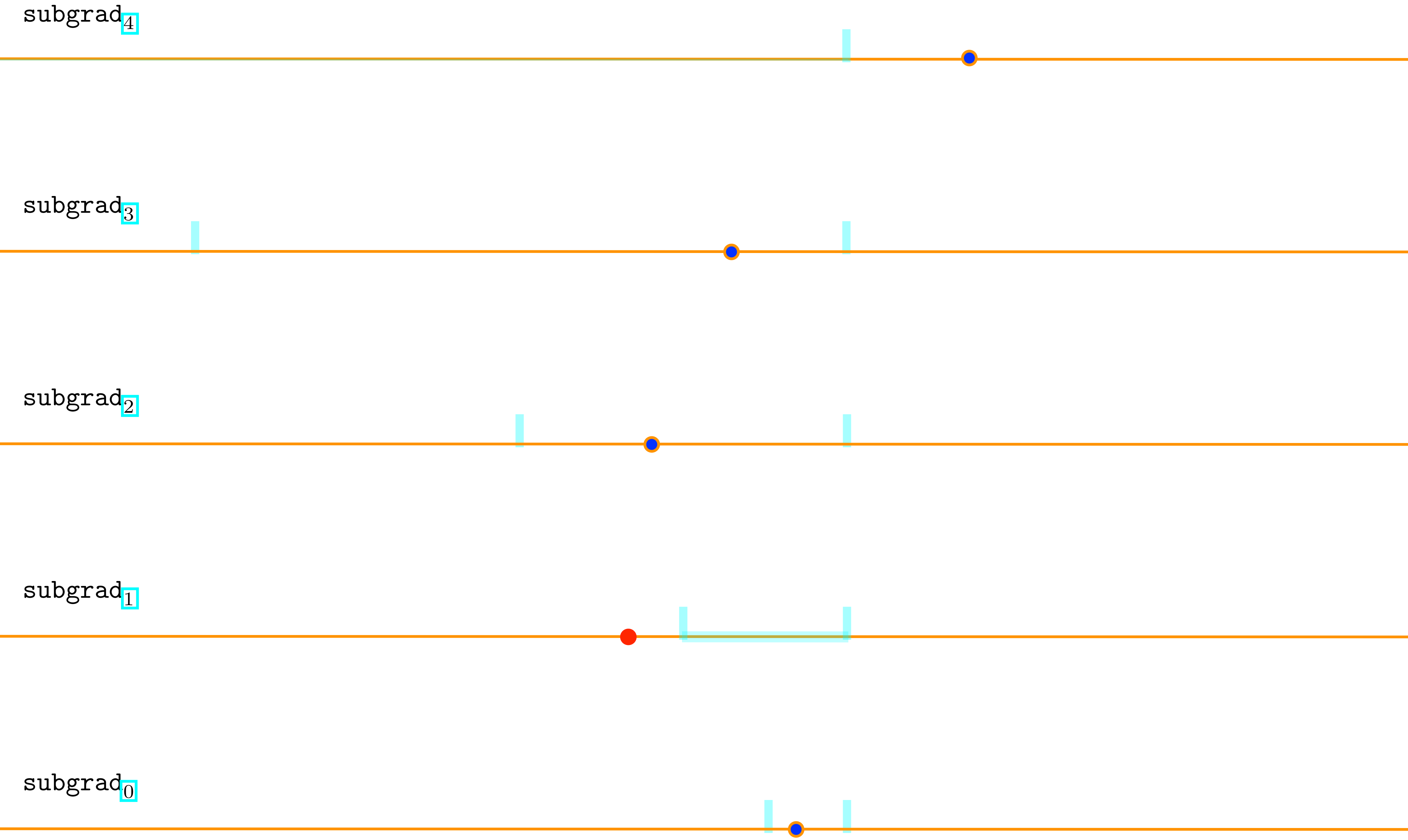
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



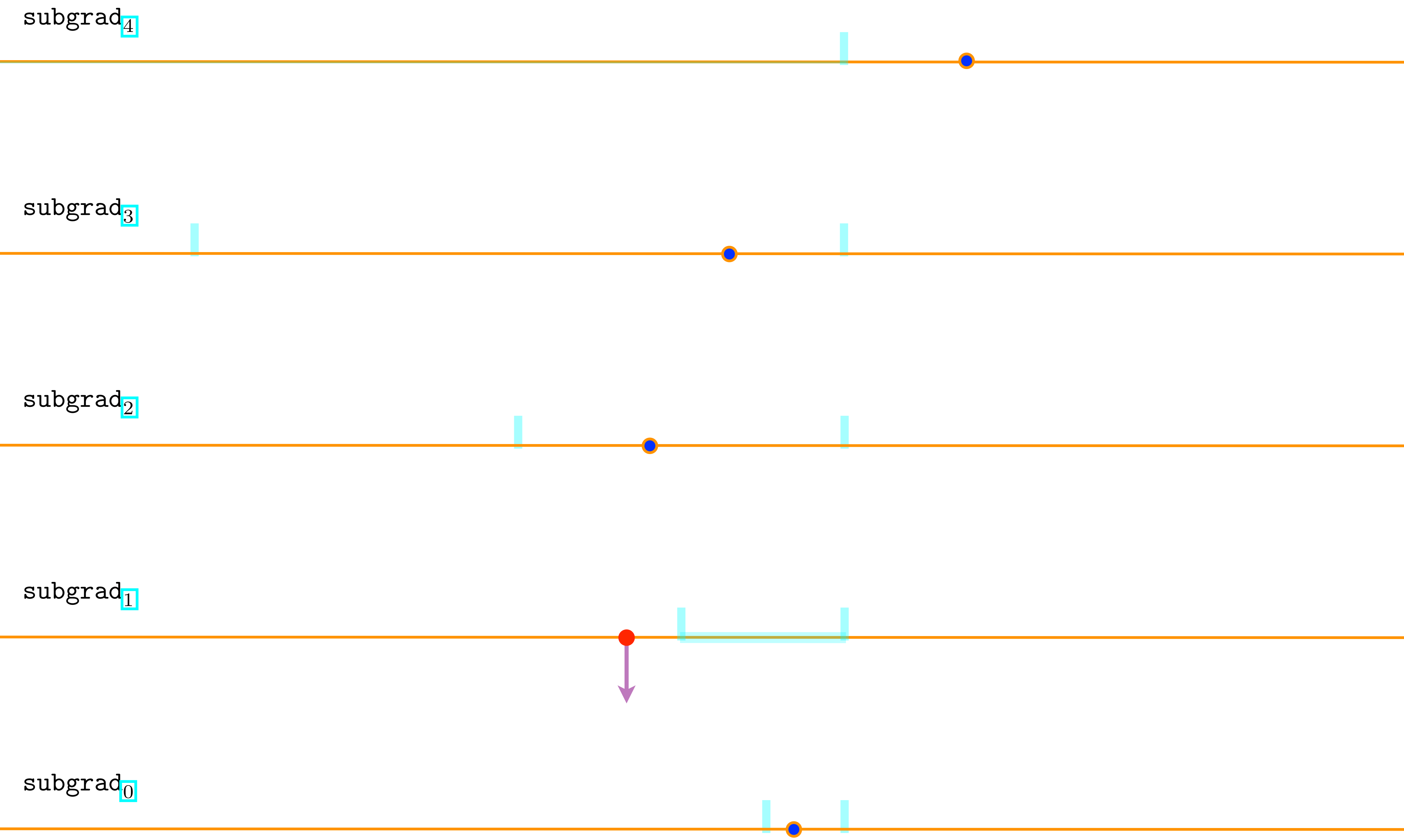
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\,x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



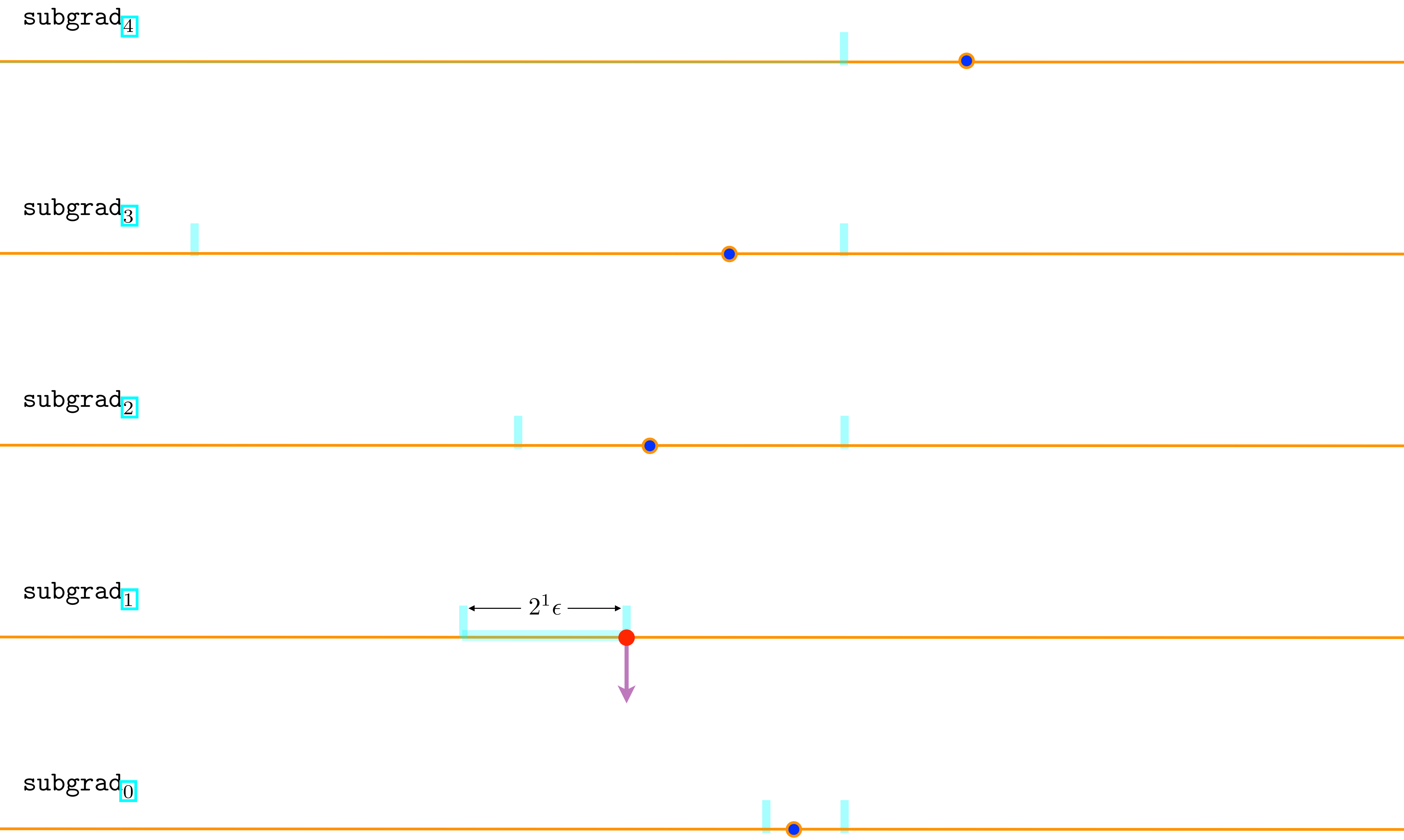
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



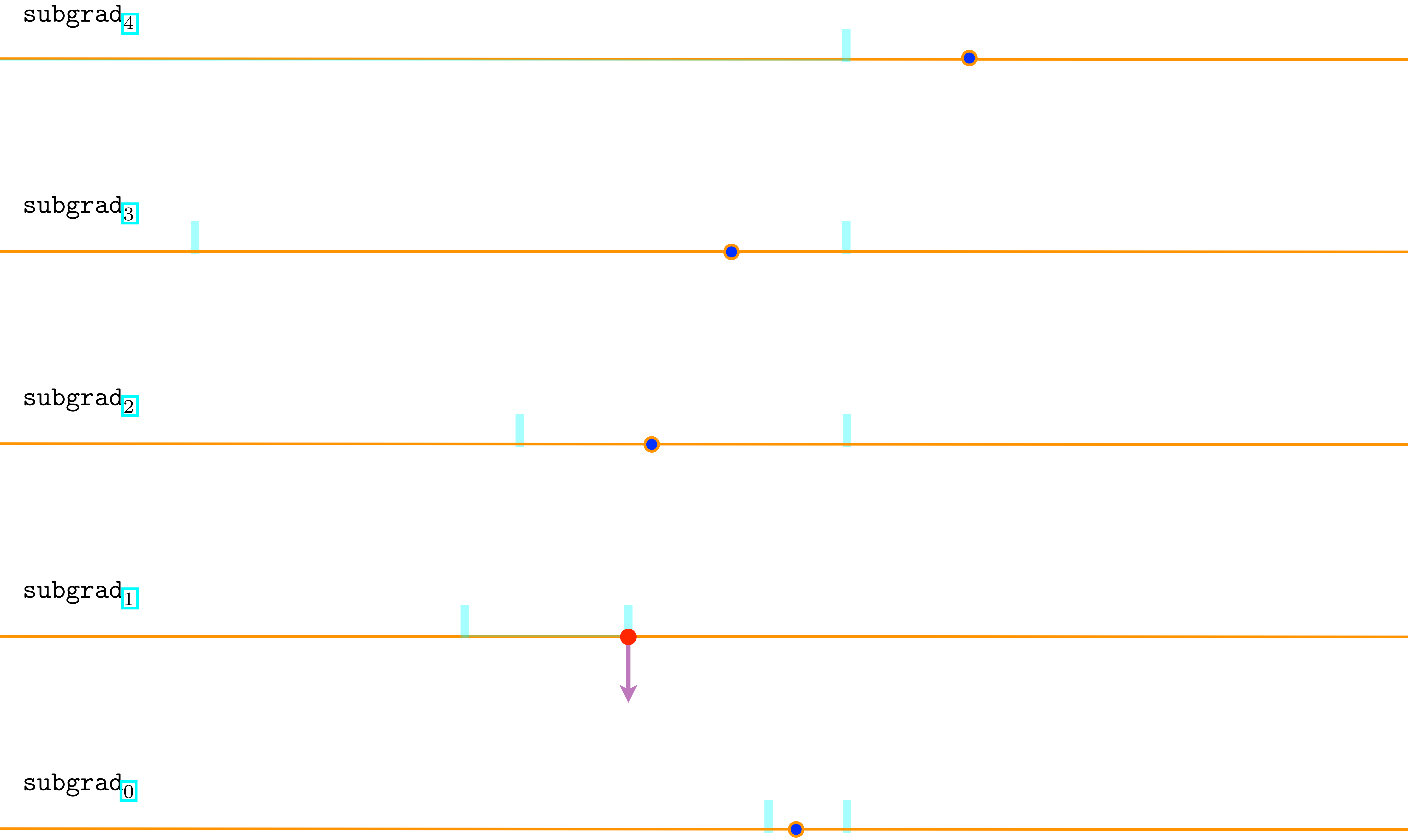
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\,x - \boxed{2^n\epsilon}\,g/\|g\|^2\,)$



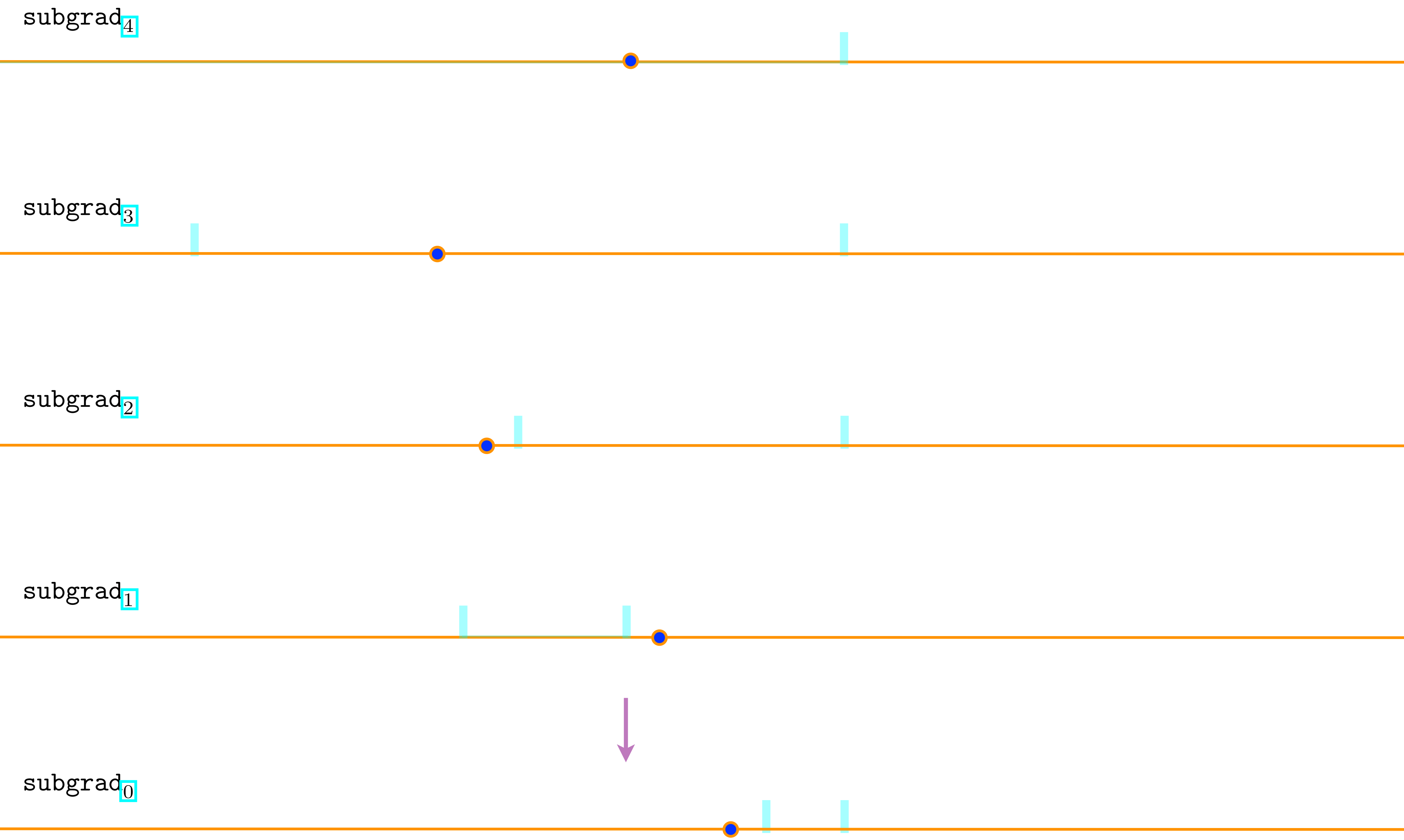
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\,x - \boxed{2^n\epsilon}\,g/\|g\|^2\,)$



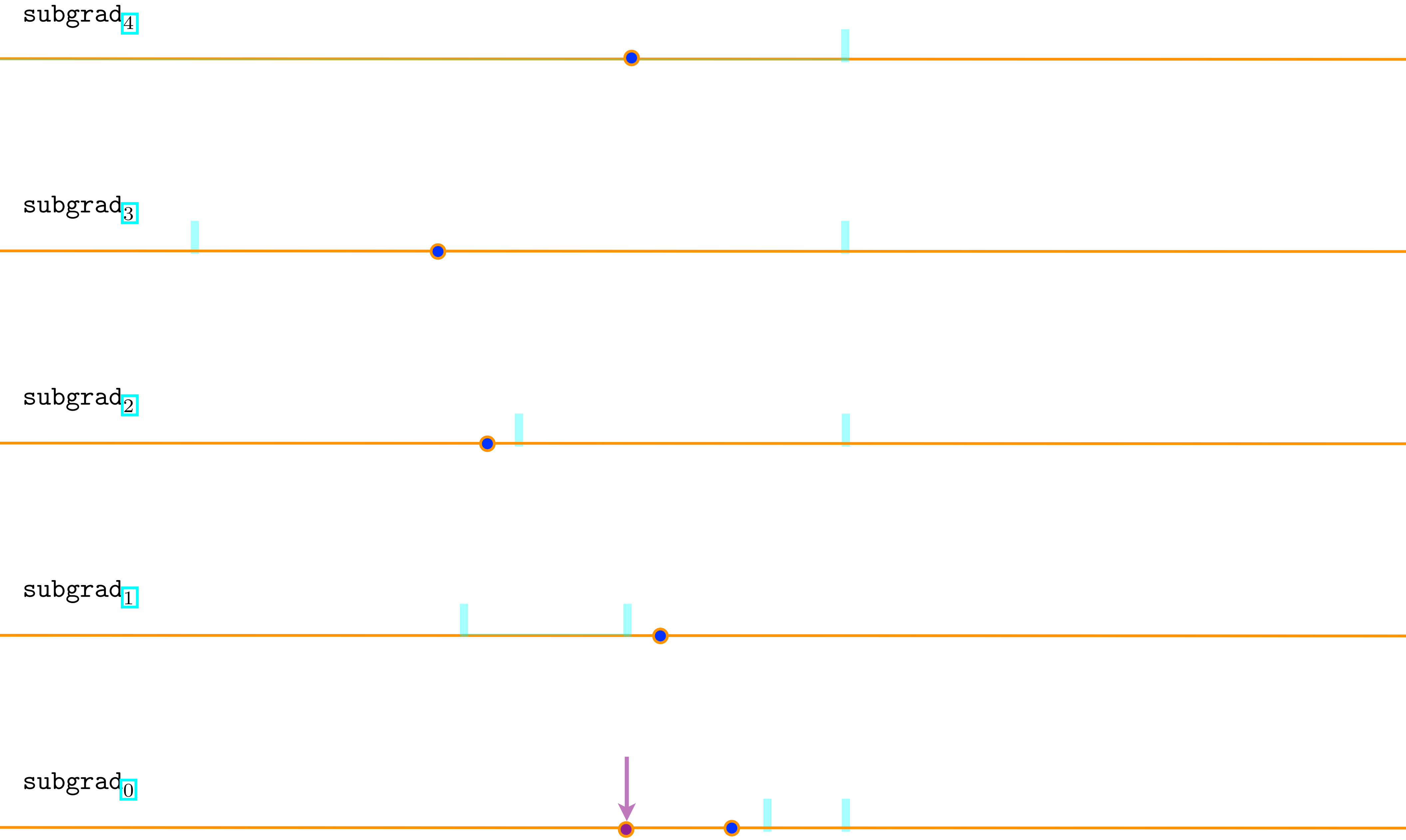
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



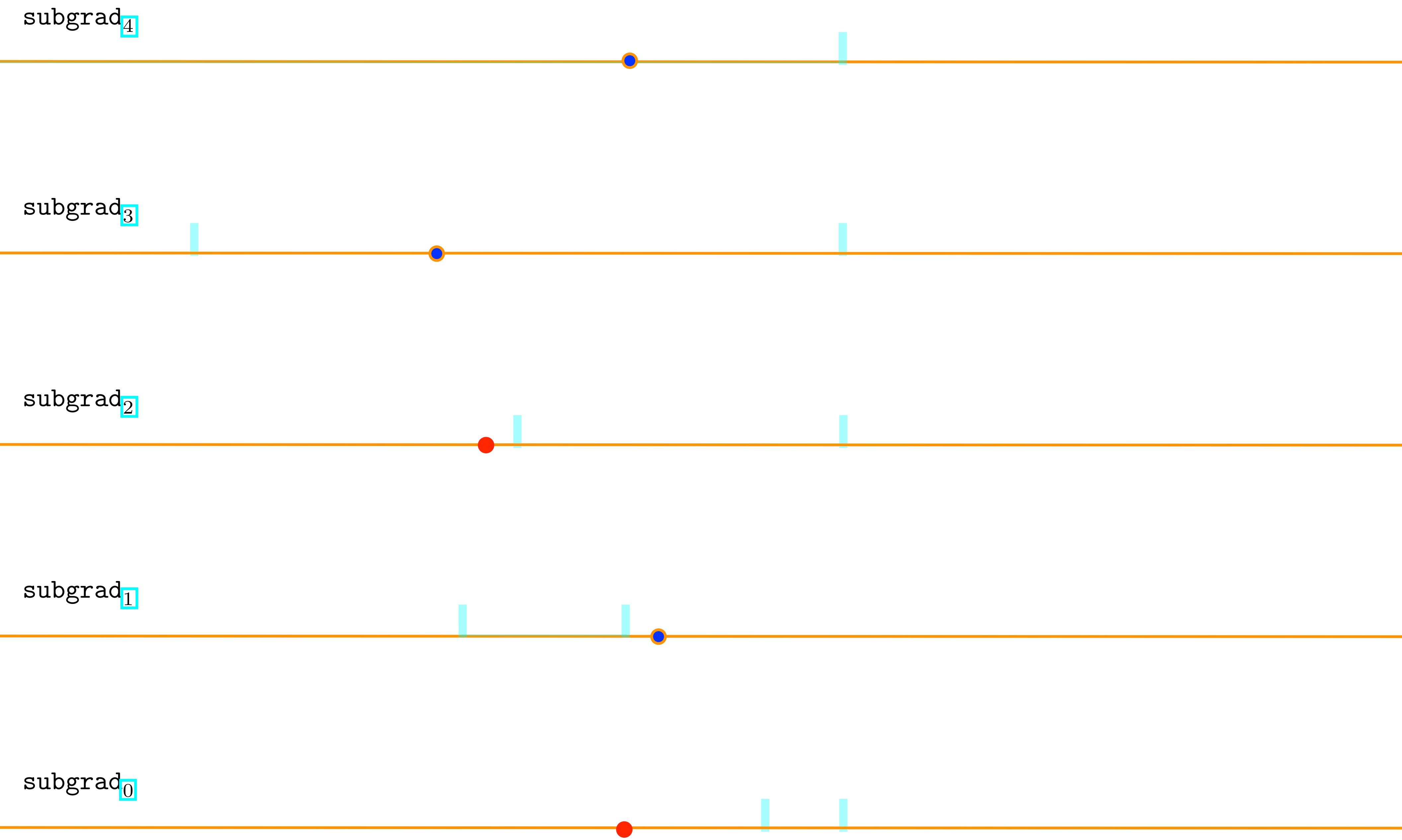
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



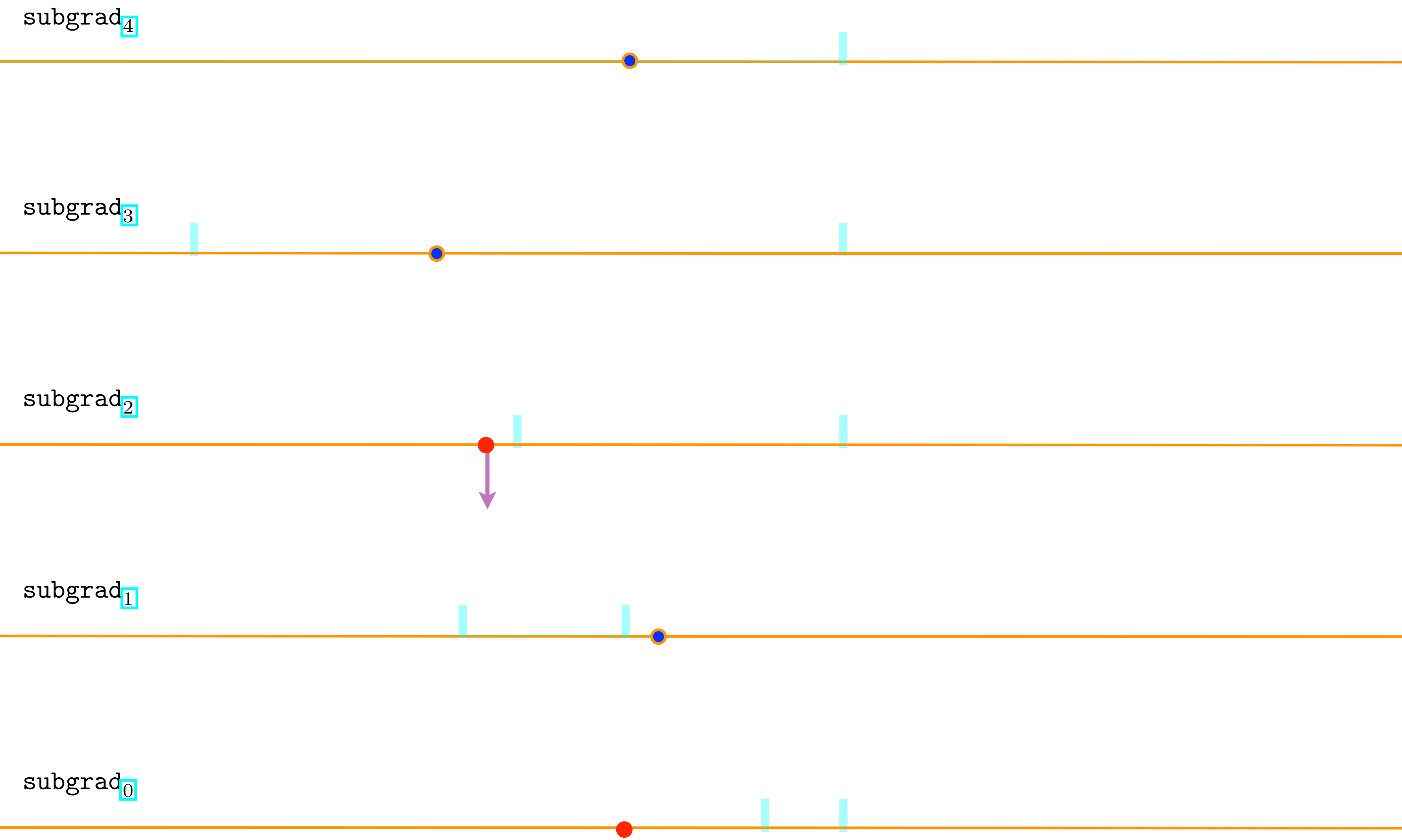
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



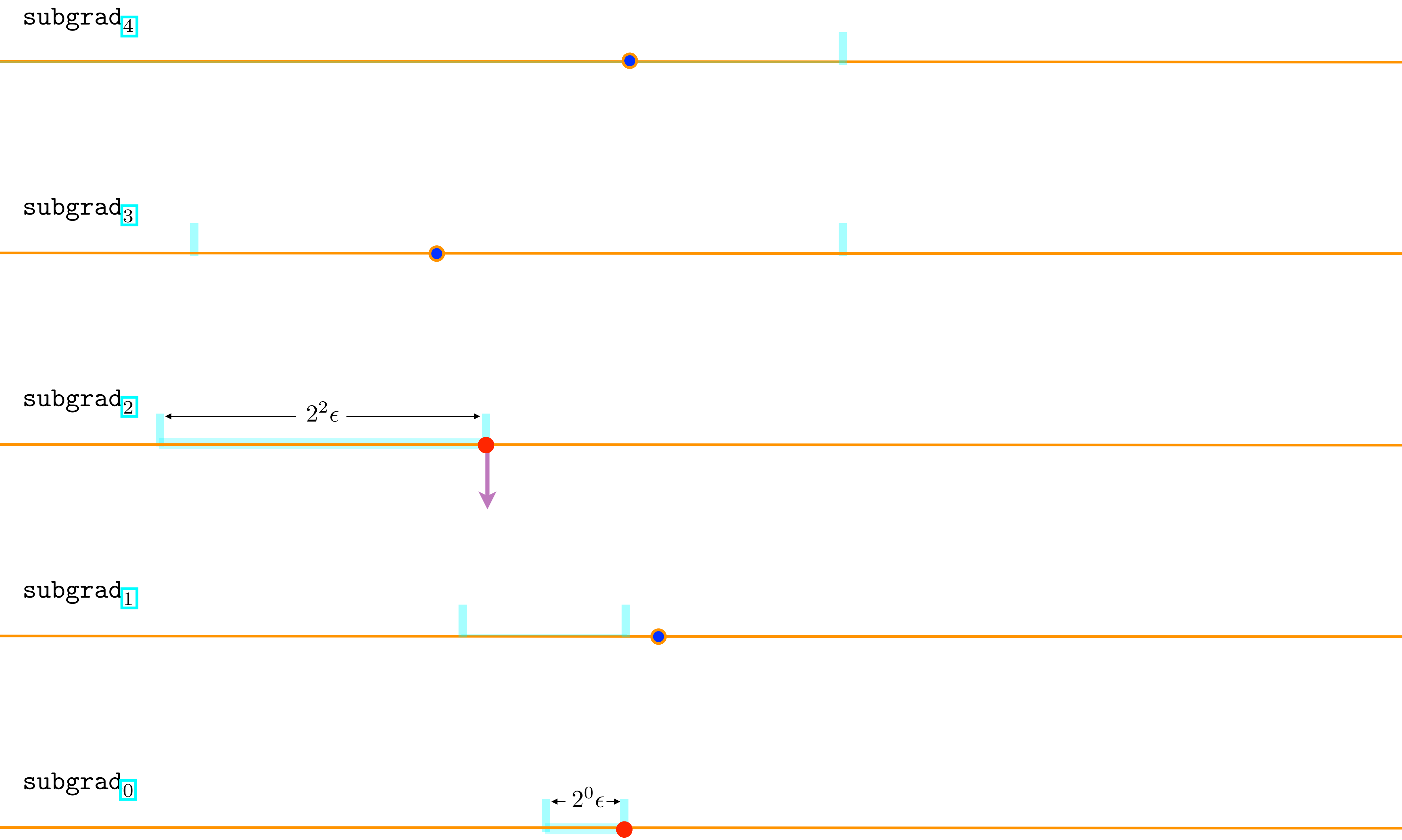
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



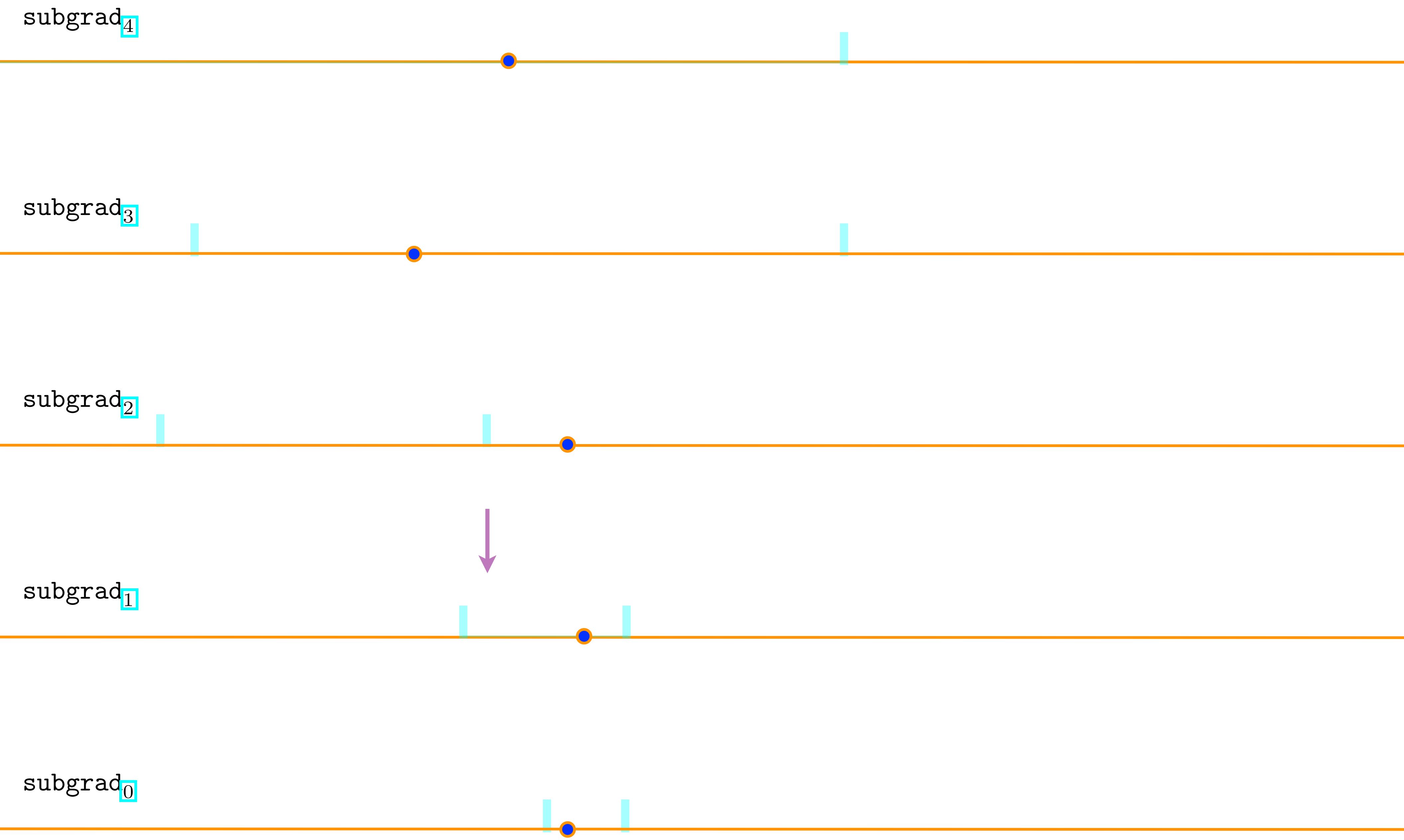
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



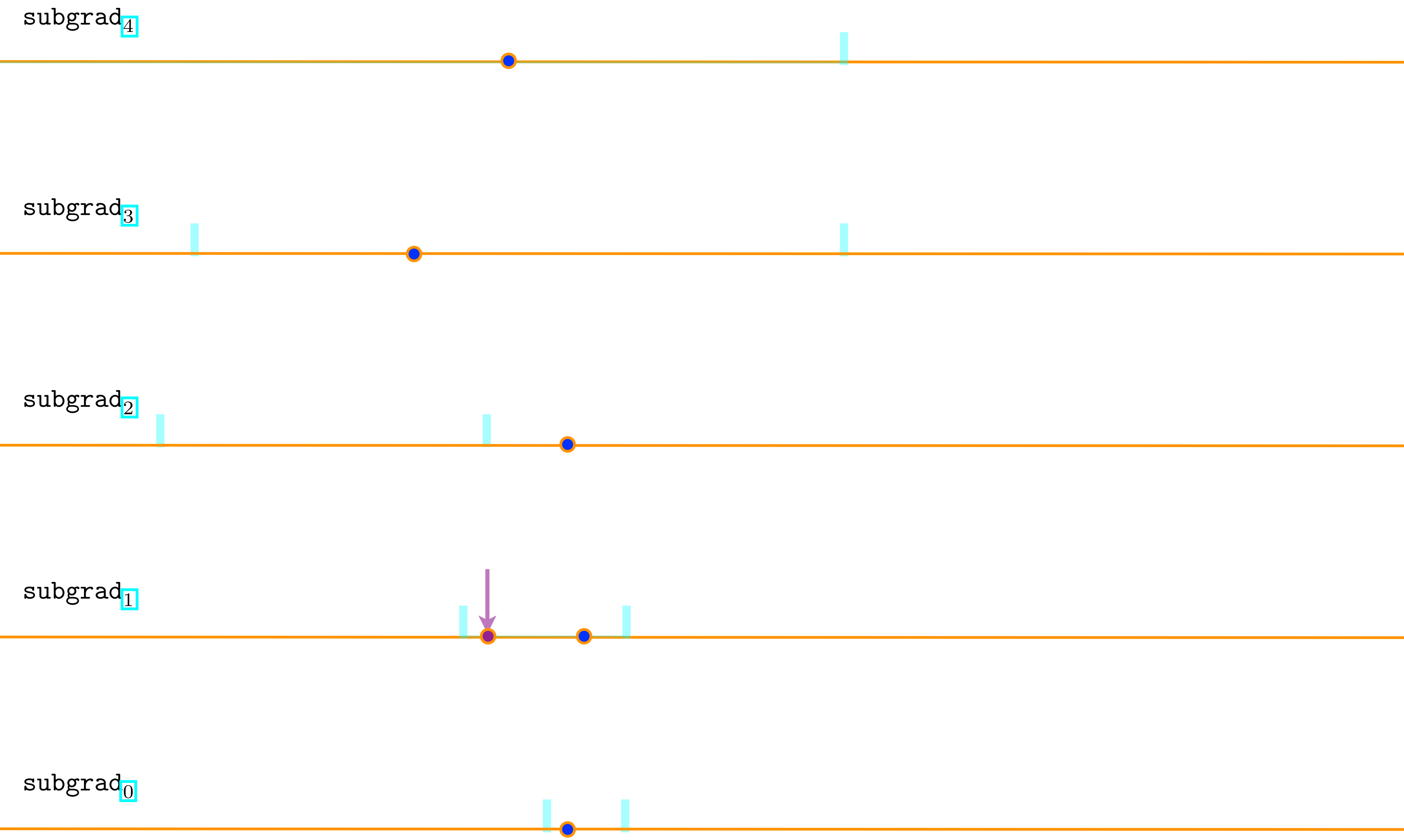
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



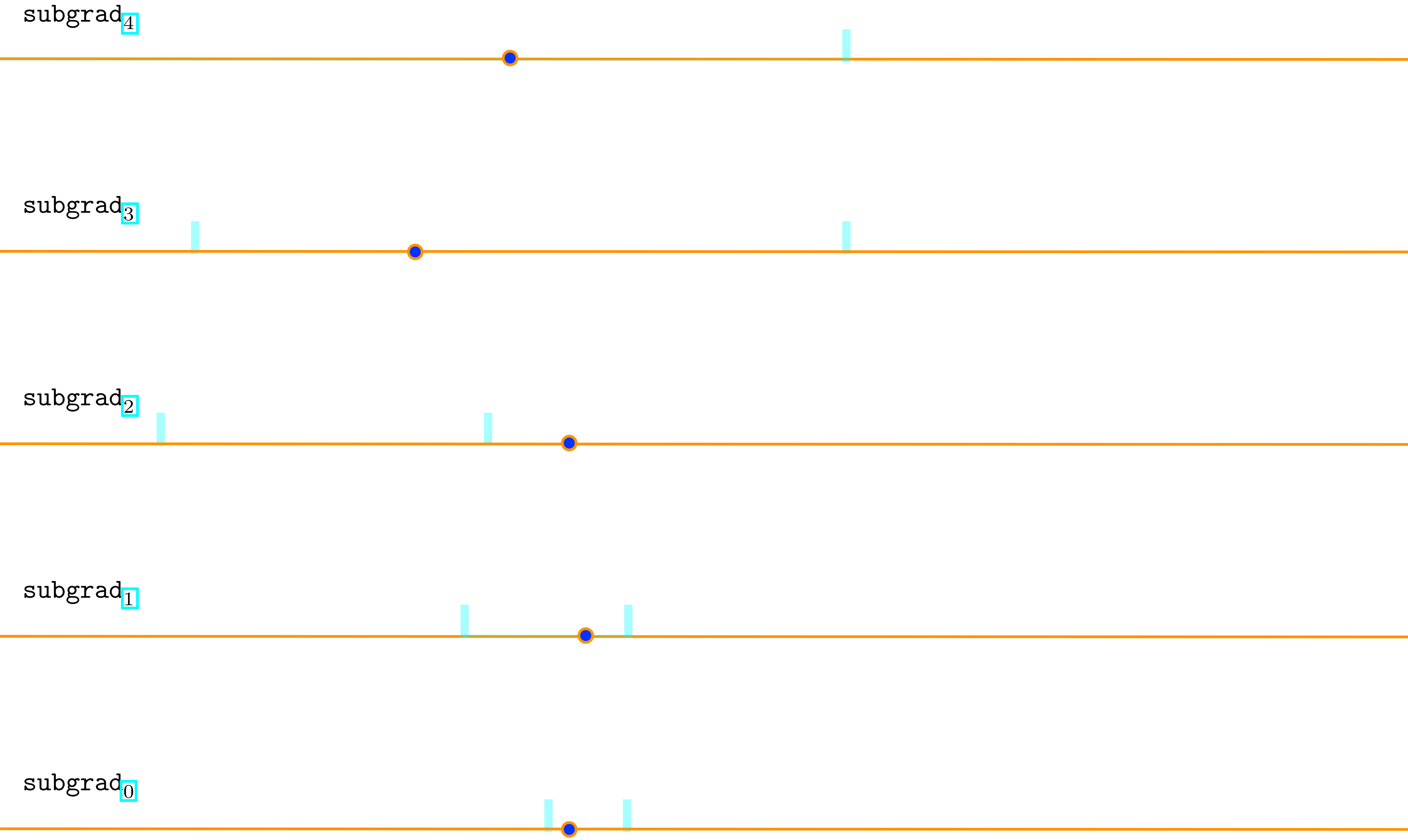
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q\left(x - \boxed{2^n \epsilon}\, g / \|g\|^2 \right)$



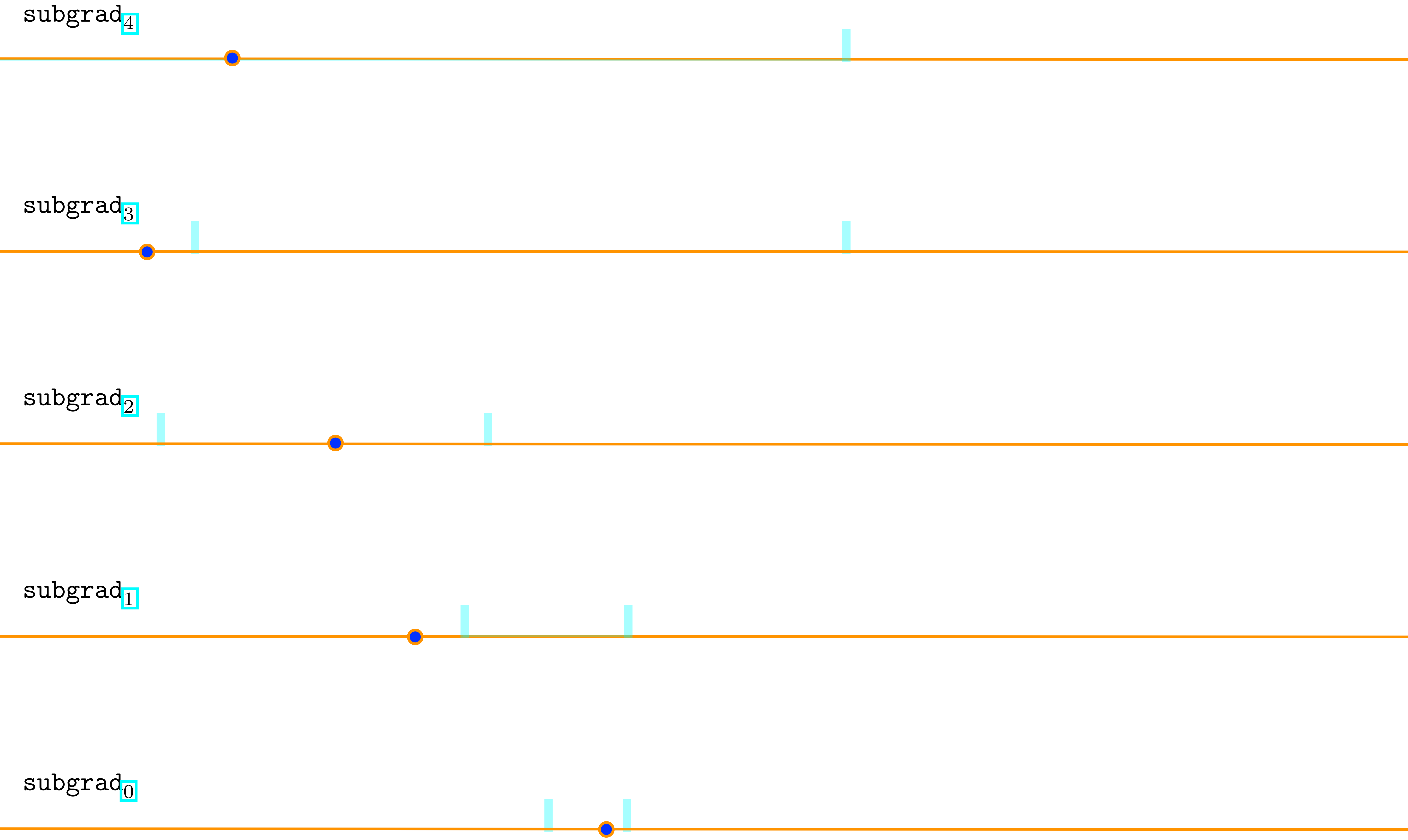
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



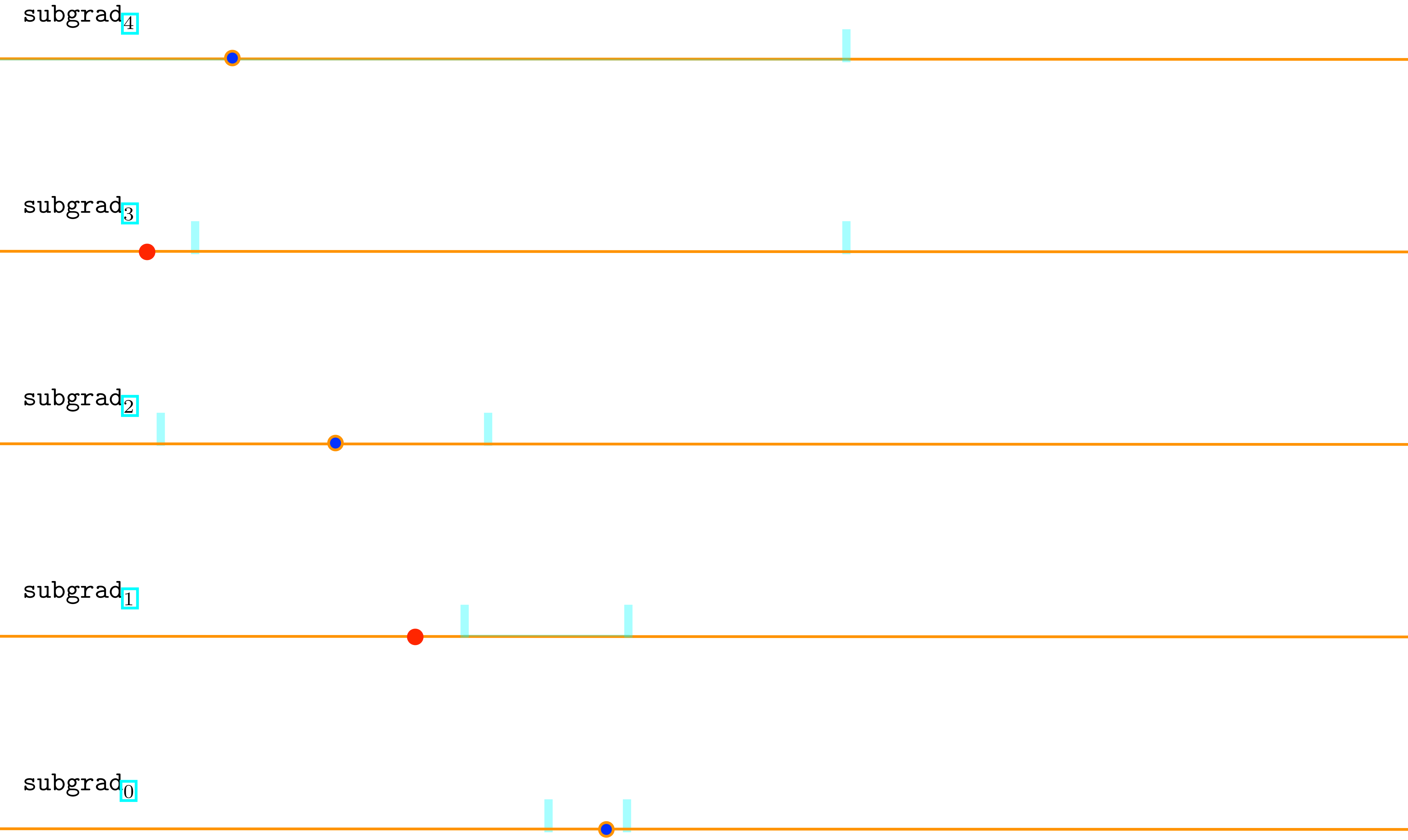
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



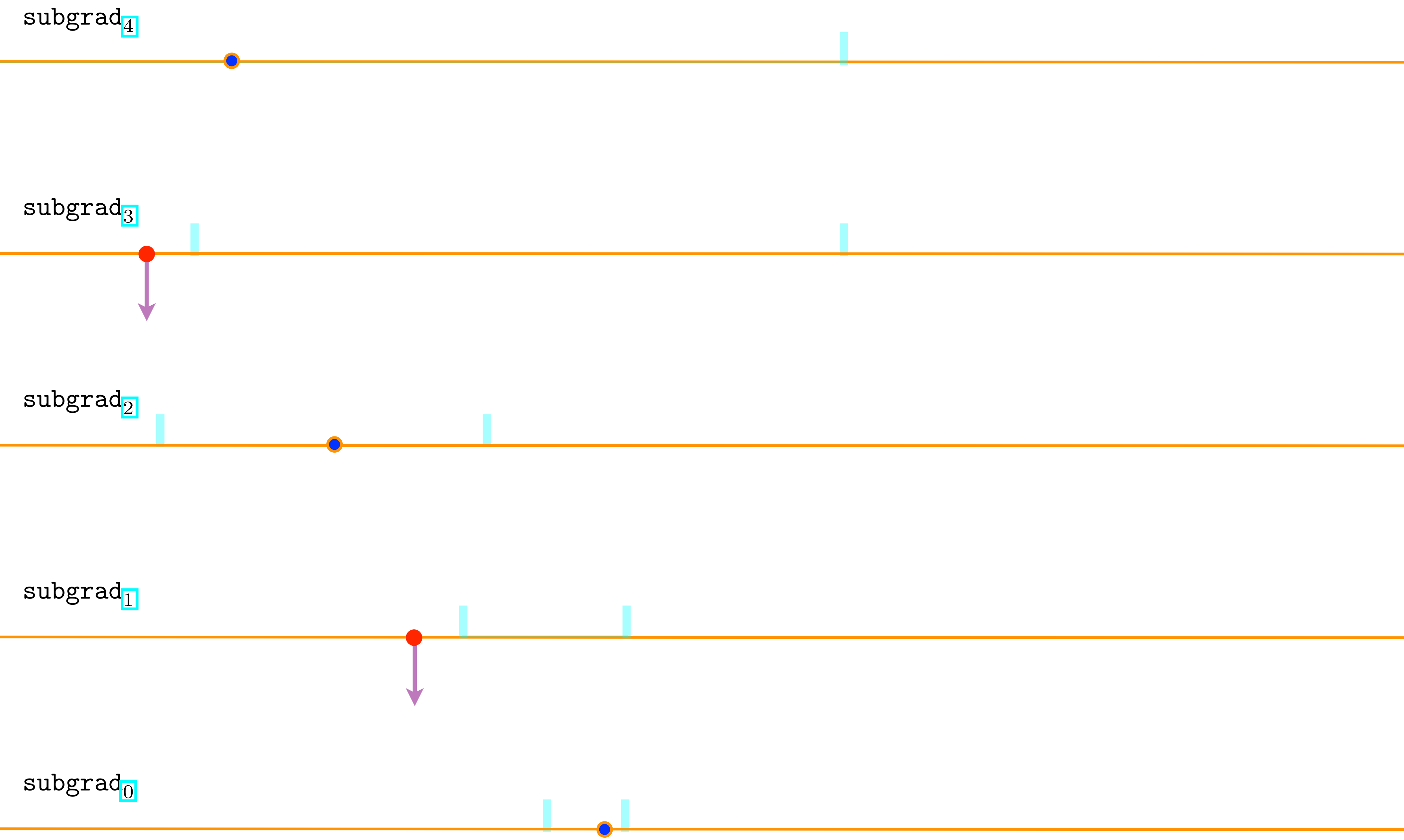
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



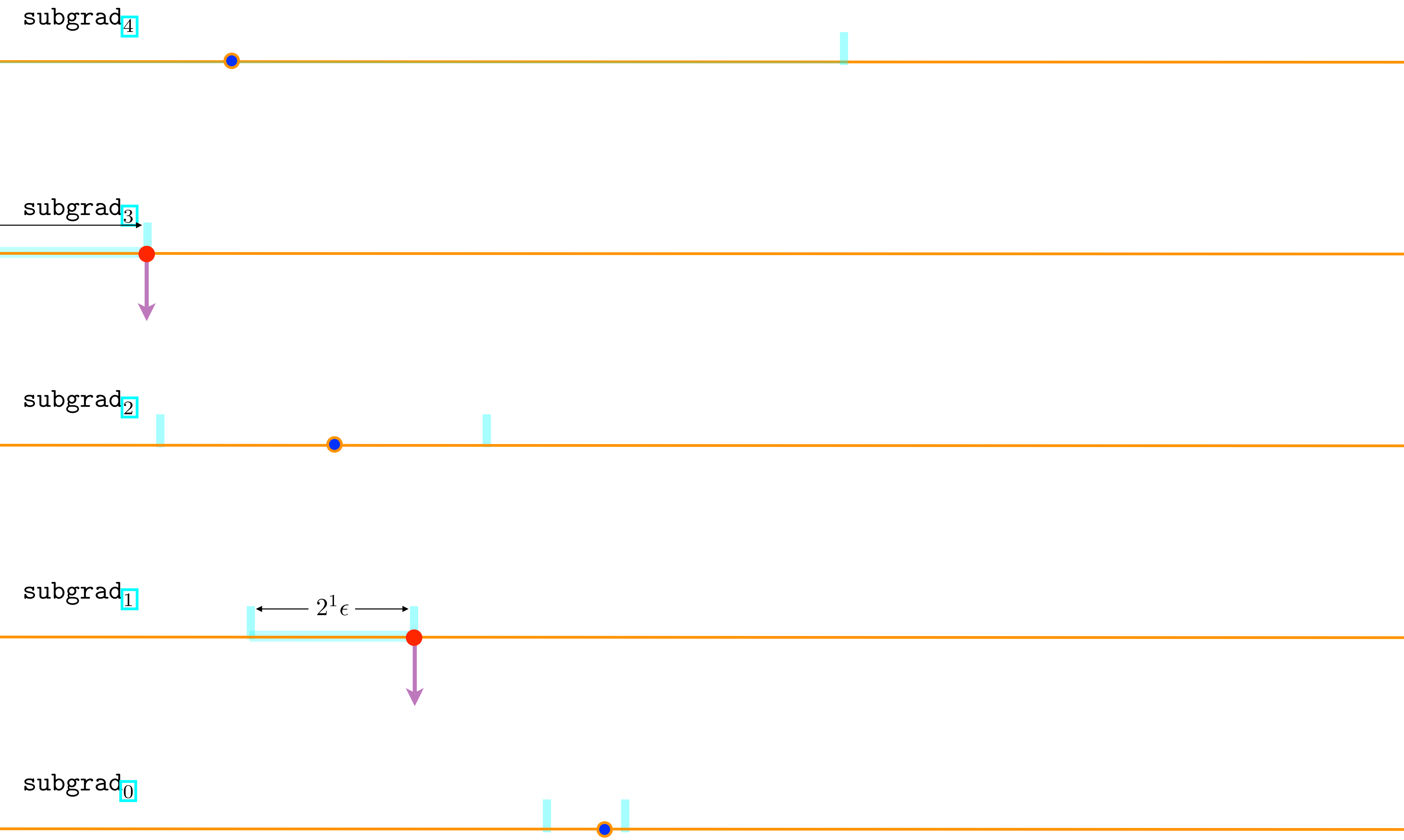
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



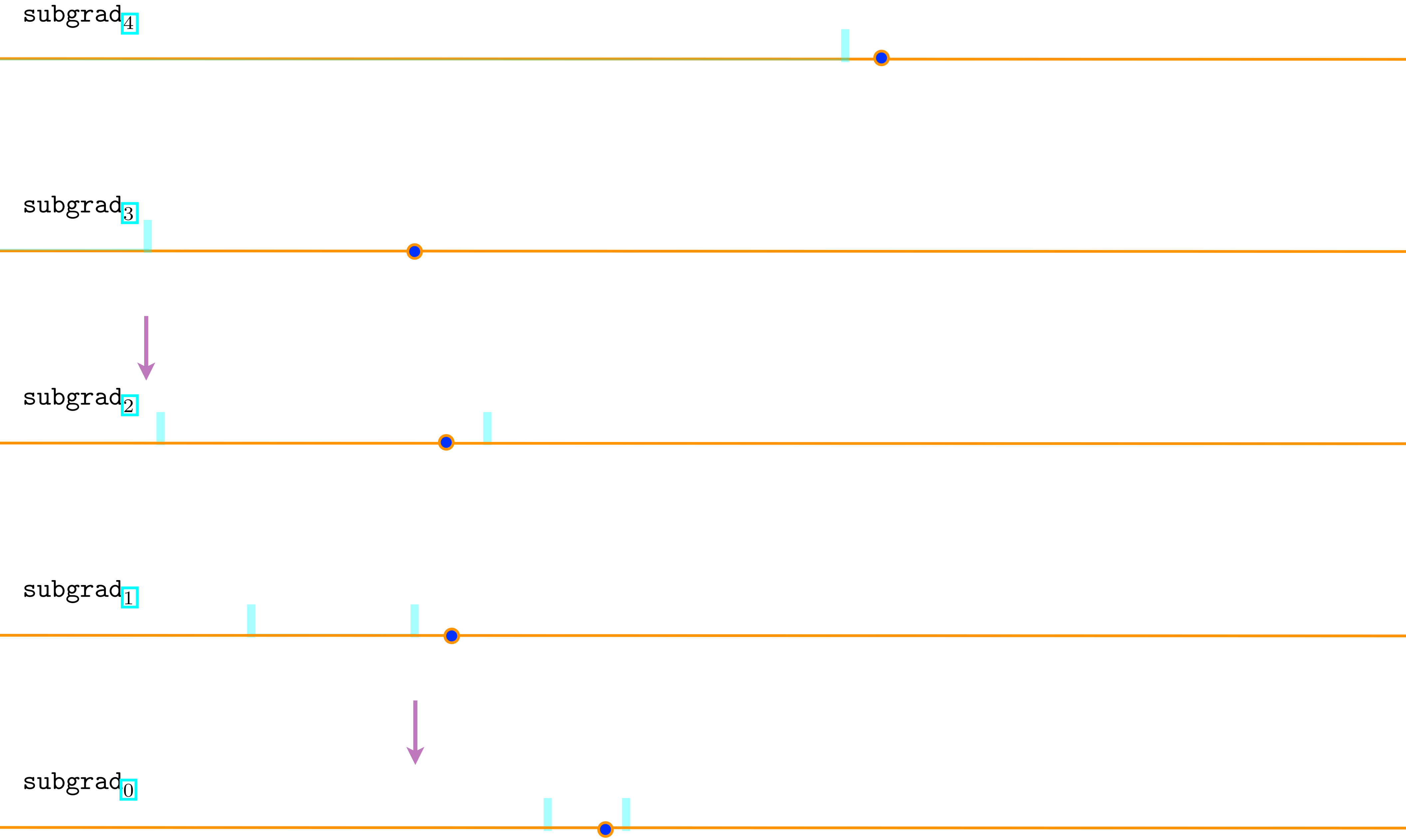
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



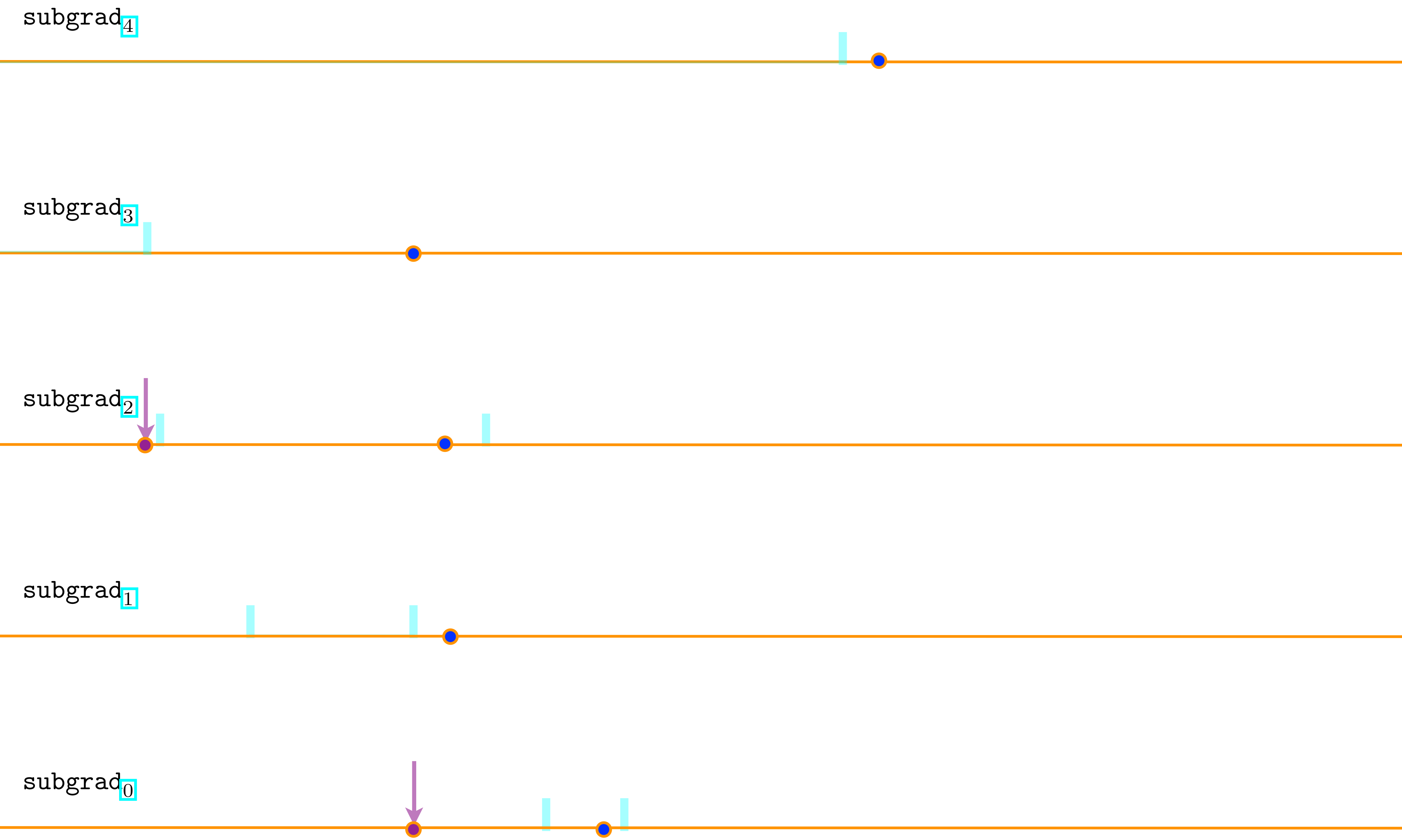
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g / \|g\|^2 \,)$



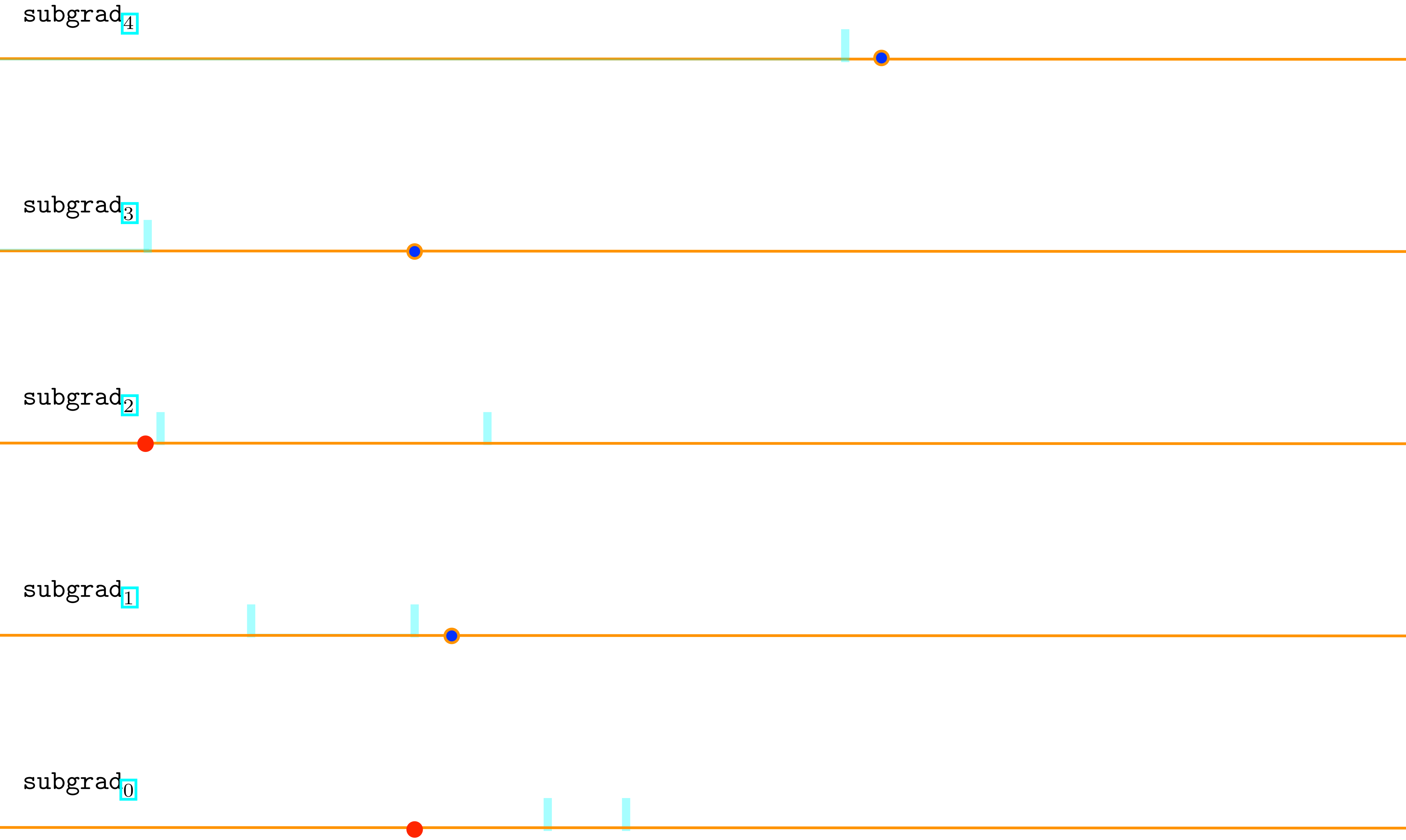
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n\epsilon}\, g/\|g\|^2\,)$



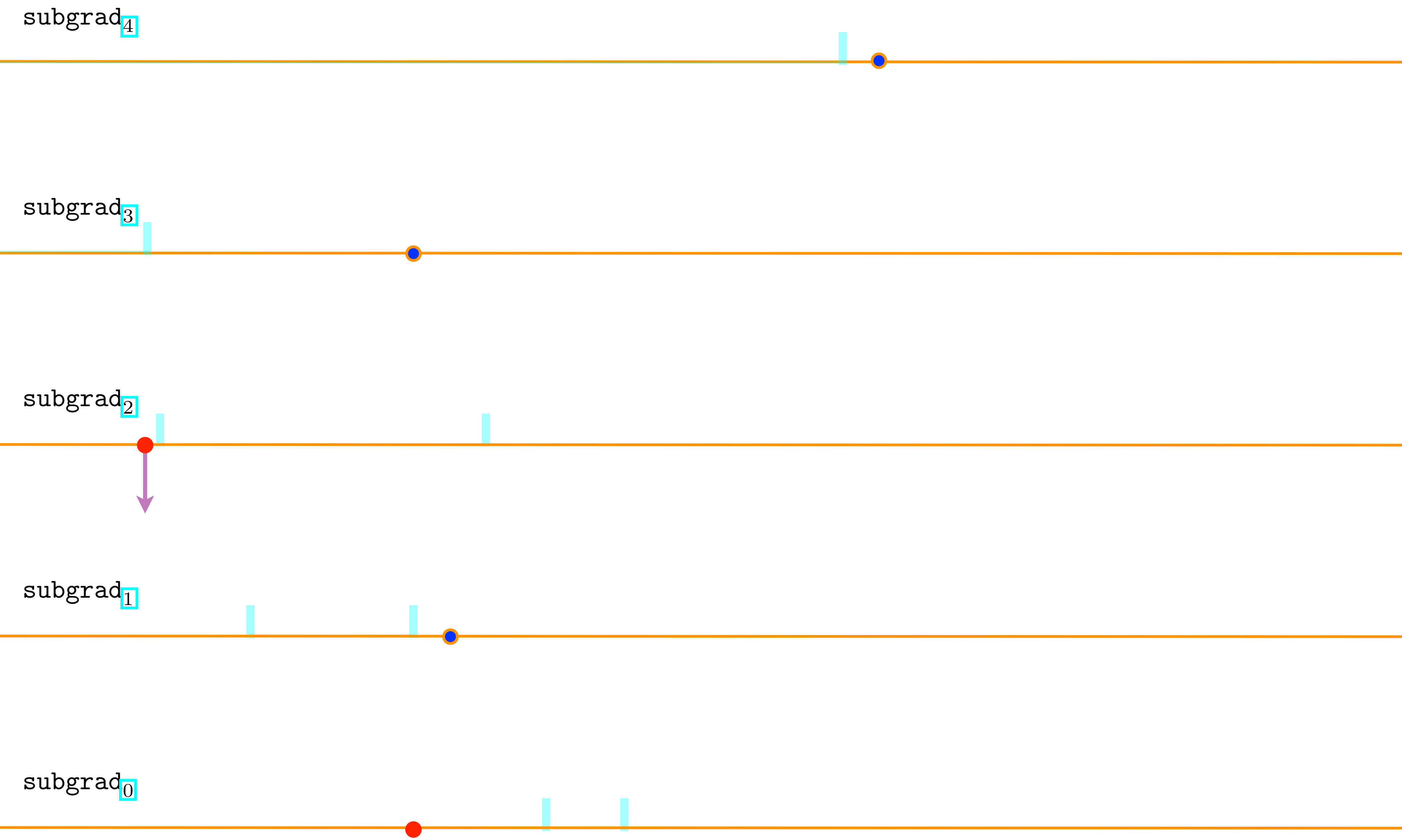
$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2 \,)$



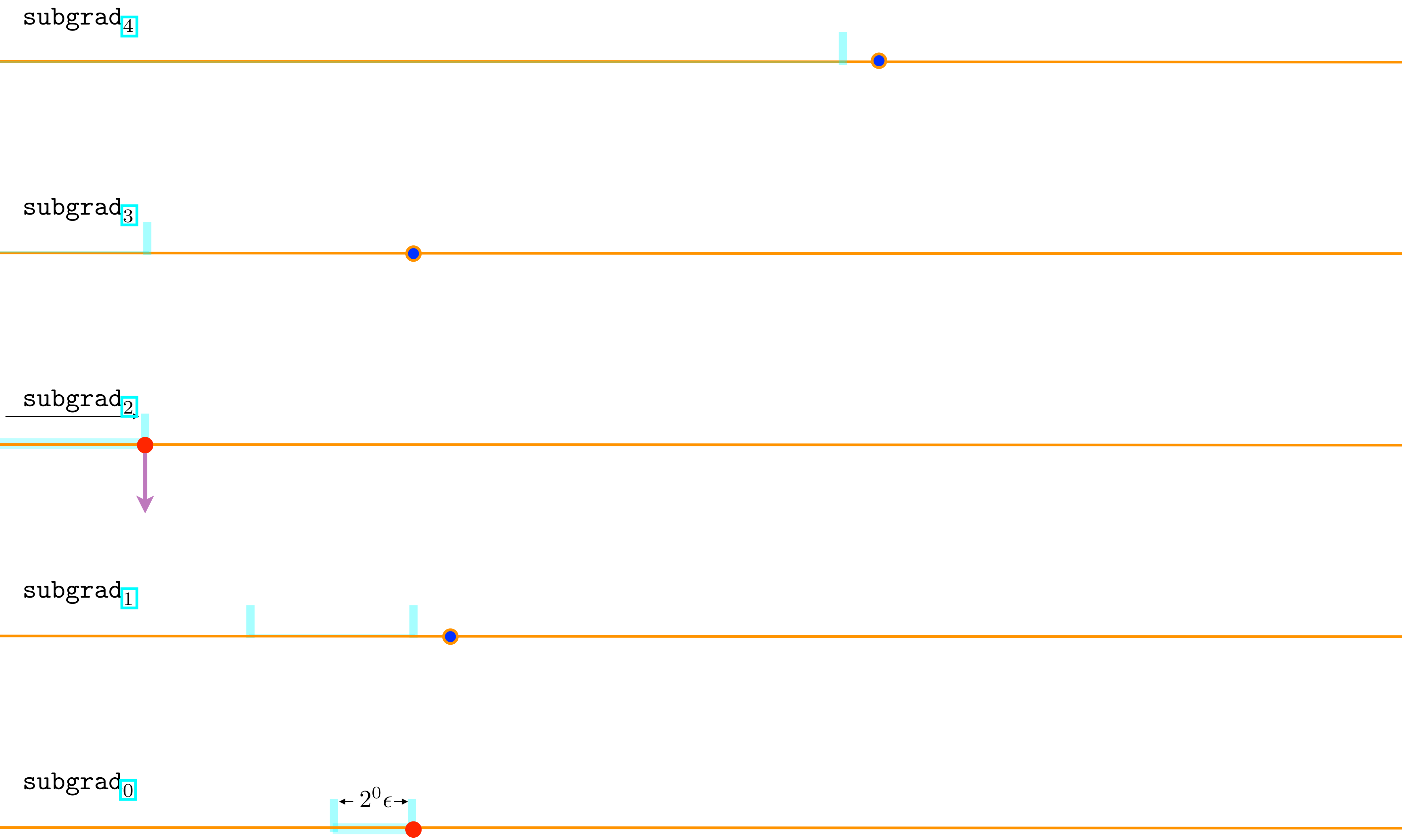
$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



$\texttt{subgrad}_{\boxed{n}} \qquad x_+ = P_Q(\, x - \boxed{2^n \epsilon}\, g/\|g\|^2\,)$



$\texttt{subgrad}_{\boxed{n}}$ $\quad x_+ = P_Q(\, x - \boxed{2^n\epsilon}\, g/\|g\|^2 \,)$



convex function

$$\min \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

closed convex set

"step size"

$$x_{k+1} = P_Q(x_k - \alpha_k g_k)$$

orthogonal projection onto Q

gradient (or subgradient) of f at x_k

Now assume f possesses quadratic growth:

$$x \text{ feasible and } \operatorname{dist}(x, X^*) \le \delta \quad \Rightarrow \quad f(x) - f^* \ge \mu \operatorname{dist}(x, X^*)^2$$

Then for the parallel scheme, $O\left(\frac{1}{\epsilon}\log_2\left(\frac{1}{\epsilon}\right)\right)$ subgradient evaluations suffice

to compute $x \in Q$ satisfying $f(x) - f^* \le \epsilon$

hiding everything besides ϵ in the big-O

By applying the parallel scheme to solving the eigenvalue optimization problem, it is possible to devise a method which within

$$O\left(\frac{1}{\epsilon}\log\left(\frac{1}{\epsilon}\right)\right) \text{ subgradient evaluations,}$$

computes a feasible symmetric matrix x_k satisfying $\dfrac{\langle c, \pi(x_k)\rangle - z^*}{\langle c, e\rangle - z^*} \leq \epsilon$

lots is being hidden in the big-O

here the big-O is relatively nice, and applies to all SDP's with bounded level sets

Recall, by contrast, the smoothing approach requires $O\left(\frac{1}{\epsilon}\right)$ gradient evaluations

But a subgradient evaluation requires only computing an eigenvector for $\lambda_{\min}(x)$, whereas in the smoothed setting, a gradient requires a full eigendecomposition of x.

Which approach is "best"? It's not clear. However, the answer is clear in the special case of linear programs.

$$\begin{array}{ll} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x\rangle = z \end{array}$$

For every polyhedral cone $\mathcal{K}$, the convex function $x \mapsto \lambda_{\min}(x)$ is piecewise linear and thus possesses "linear growth":

$$x \text{ feasible} \quad \Rightarrow \quad \lambda_{\min}(x_z^*) - \lambda_{\min}(x) \geq \mu_z \text{dist}(x, X_z^*)$$

For the parallel scheme, $O\left(\log\left(\frac{1}{\epsilon}\right)^2\right)$ subgradient evaluations suffice

$$
\begin{array}{rl} \min & \langle c, x\rangle \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}
\qquad \equiv \qquad
\begin{array}{rl} \max & \lambda_{\min}(x) \\ \text{s.t.} & Ax = b \\ & \langle c, x\rangle = z \end{array}
$$

"Efficient" subgradient methods for general convex optimization

Accelerated first-order methods for hyperbolic programming

with Ben Grimmer:

A simple nearly-optimal restart scheme for speeding-up first order methods

---

Ben Grimmer: *Radial subgradient method*