3.6 The least-squares problem

Create a method for a specific problem.

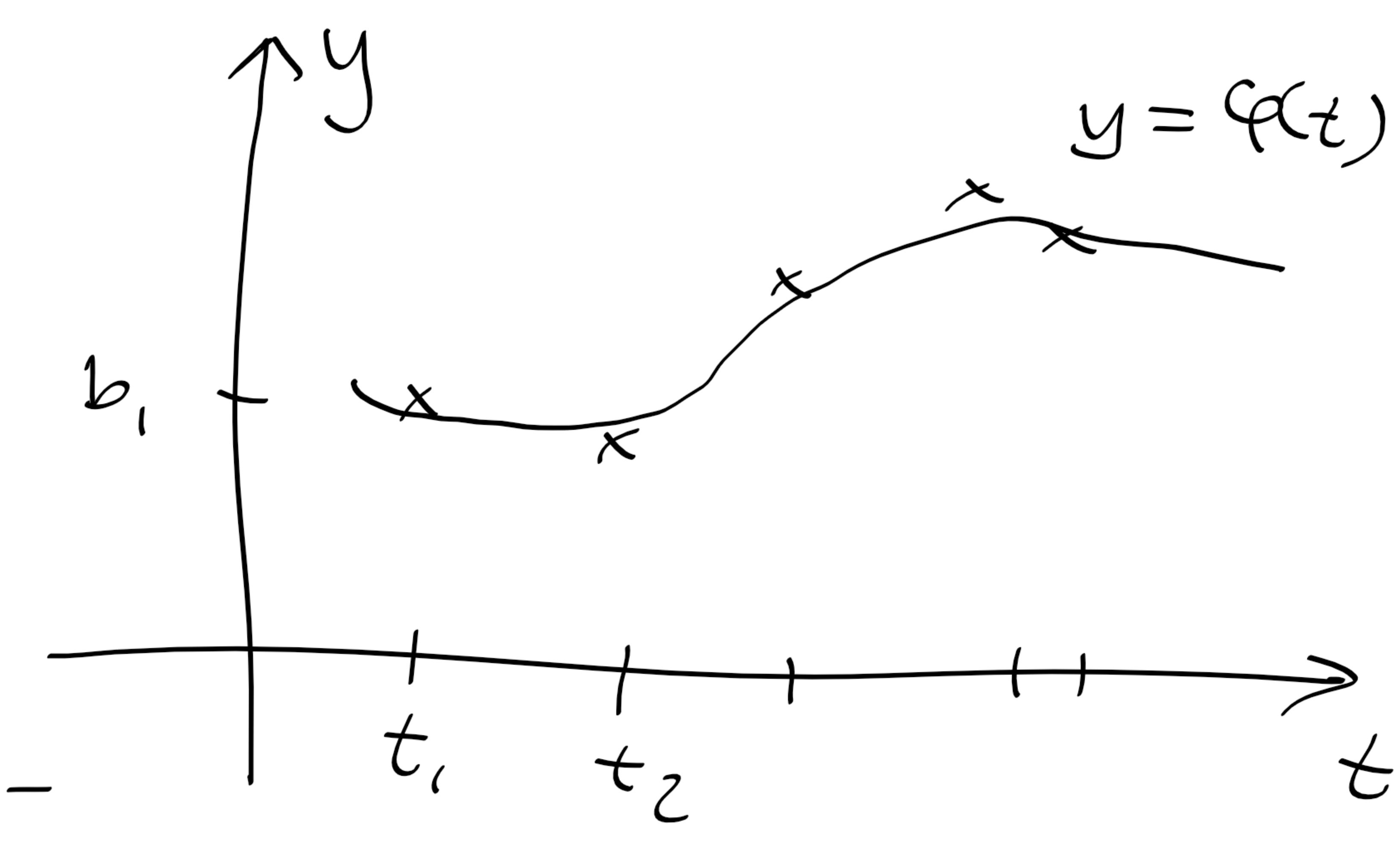
Curve fitting:

Given data points

(t_i, b_i), $i = 1, \dots, m$

fit a curve $y = \varphi(t)$

where φ contains n parameters



$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{by minimizing}$$

$$f(x) = \sum_{i=1}^{m} \big(\underbrace{\varphi(t_i; x) - b_i}_{r_i(x) \text{ residual}} \big)^2 = r(x)^T r(x) = \| r(x) \|^2$$

The linear case

$\varphi(t; \cdot)$ is linear, e.g.

$$\varphi(t; x) = x_1 t + x_2 t^2 + \dots + x_n t^n$$

or $\varphi(t; x) = x_1 \alpha_1(t) + \dots + x_n \alpha_n(t)$

with $\alpha_i(t)$ any (nonlinear) function

$$\begin{cases} x_1 \alpha_1(t_1) + \dots + x_n \alpha_n(t_1) = b_1 \\ \vdots \\ x_1 \alpha_1(t_m) + \dots + x_n \alpha_n(t_m) = b_m \end{cases} \iff \underset{m \times n}{A} x = b$$

overdetermined system when $m > n$

The residual vector is $r(x) = Ax - b$

$f(x) = \|r(x)\|^2 = \|Ax-b\|^2 = (Ax-b)^T(Ax-b)$

$= x^TA^TAx - x^TA^Tb - \underbrace{b^TAx}_{(b^TAx)^T = x^TA^Tb} + b^Tb$

$= \frac{1}{2}x^T(2A^TA)x - 2x^TA^Tb + b^Tb$

Quadratic function with Hessian $\underset{(n\times n)}{2A^TA}$.

Minimum when

$\nabla f(x) = 0 \iff 2A^TAx - 2A^Tb = 0$

$\iff \underline{A^TAx = A^Tb}$ normal equations

$x = \underbrace{(A^TA)^{-1}A^T}b$

pseudoinvers of A $(m\times n, m>n)$

if A^TA is invertible

- A^TA is pos. semidef. because $x^TA^TAx = (Ax)^TAx = \|Ax\|^2 \geq 0$
- If $\operatorname{rank} A = n \iff$ columns of A are linearly independent $\iff$ $[Ax = 0 \iff x = 0]$, then A^TA is pos. def.; hence invertible.

The nonlinear case (Gauss-Newton)

$$f(x) = \sum_{i=1}^{m} r_i(x)^2 \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\nabla f(x) = \sum 2 r_i(x) \nabla r_i(x) = 2 \left(\nabla r_1 \ \dots \ \nabla r_m\right) \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}$$

$$= 2\, J(x)^T r(x)$$

with the Jacobian $J(x) = \begin{pmatrix} \nabla r_1^T \\ \vdots \\ \nabla r_m^T \end{pmatrix}$

$$\nabla^2 f(x) = \nabla (\nabla f(x))^T = 2 \nabla \sum r_i(x) \nabla r_i(x)^T$$

$$= 2 \sum \nabla r_i(x) \nabla r_i(x)^T + 2 \sum r_i(x) \nabla^2 r_i(x)$$

$$= 2 \left(\nabla r_1 \ \dots \ \nabla r_m\right) \begin{pmatrix} \nabla r_1^T \\ \vdots \\ \nabla r_m^T \end{pmatrix} + \text{———} \ '' \ \text{———}$$

$$= \underbrace{2\, J(x)^T J(x)}_{\text{first derivatives}} + \underbrace{2 \sum r_i(x) \nabla^2 r_i(x)}_{\text{second derivatives}}$$

$$\approx 2\, J(x)^T J(x) \qquad \text{for small } r_i(x)$$

Newton's method with this approximation is Gauss-Newton's method:

$$x_{k+1} = x_k - \left(J(x_k)^T J(x_k)\right)^{-1} J(x_k)^T r(x_k)$$

with line search: include λ_k

9.2 The penalty function method

Idea: convert the constrained problem

$$(P) \quad \begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in S = \left\{ x \in X : \begin{array}{l} g(x) \le 0, \\ h(x) = 0 \end{array} \right\} \end{aligned}$$

to an unconstrained one where the objective function is large when $x \notin S$:

$$\underset{x \in X}{\text{minimize}} \quad \varphi(x) = f(x) + \mu \alpha(x)$$

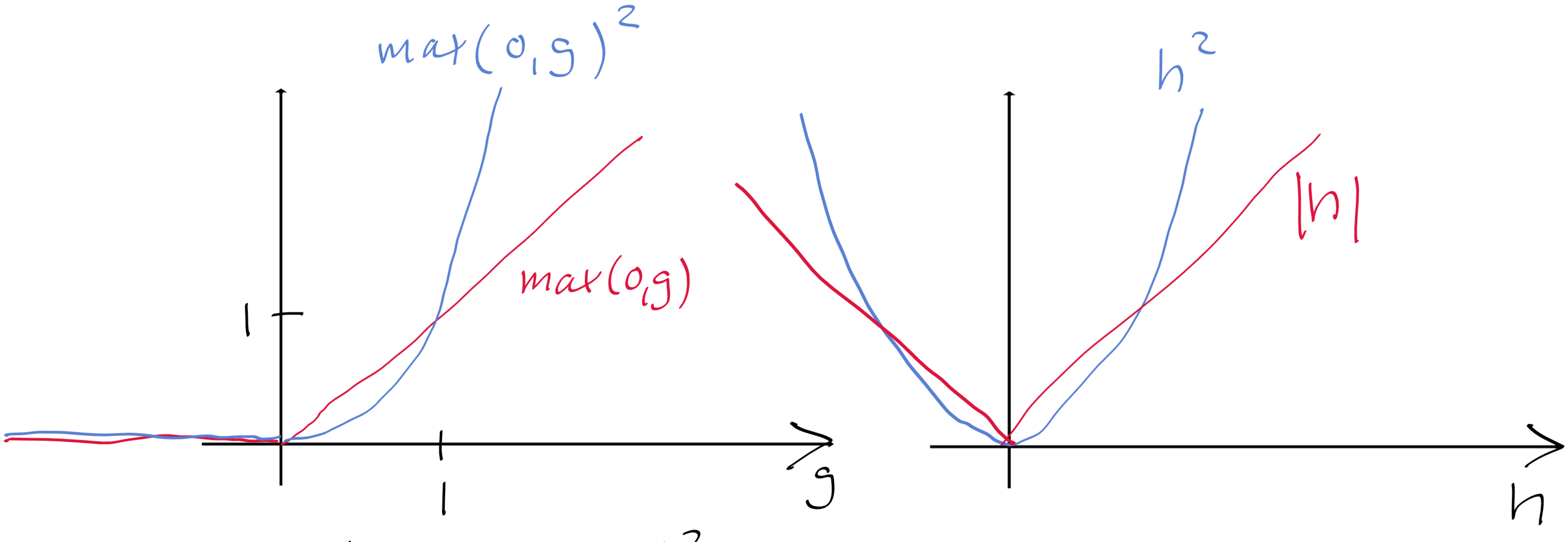
where $\mu > 0$ and the penalty function $\alpha(x)$ satisfies

- $\alpha(x) = 0$ when $x \in S$
- $\alpha(x) > 0$ when $x \notin S$
- $\begin{cases} \alpha \in C & (\text{if eg. Nelder-Mead is used}) \\ \alpha \in C^1 & (\text{if CG or Quasi-Newton is used}) \end{cases}$

Typical penalty functions:

$$\alpha_0(x) = \sum_{i=1}^{m} \max(0, g_i(x)) + \sum_{j=1}^{l} |h_j(x)| \in C$$

$$\alpha_1(x) = \sum (\max(0, g_i(x)))^2 + \sum h_j(x)^2 \in C^1$$



Note: $\frac{d}{dg} (\max(0,g))^2 = 2 \max(0,g)$

Ex. in 1D: $f(x) = \ln x$

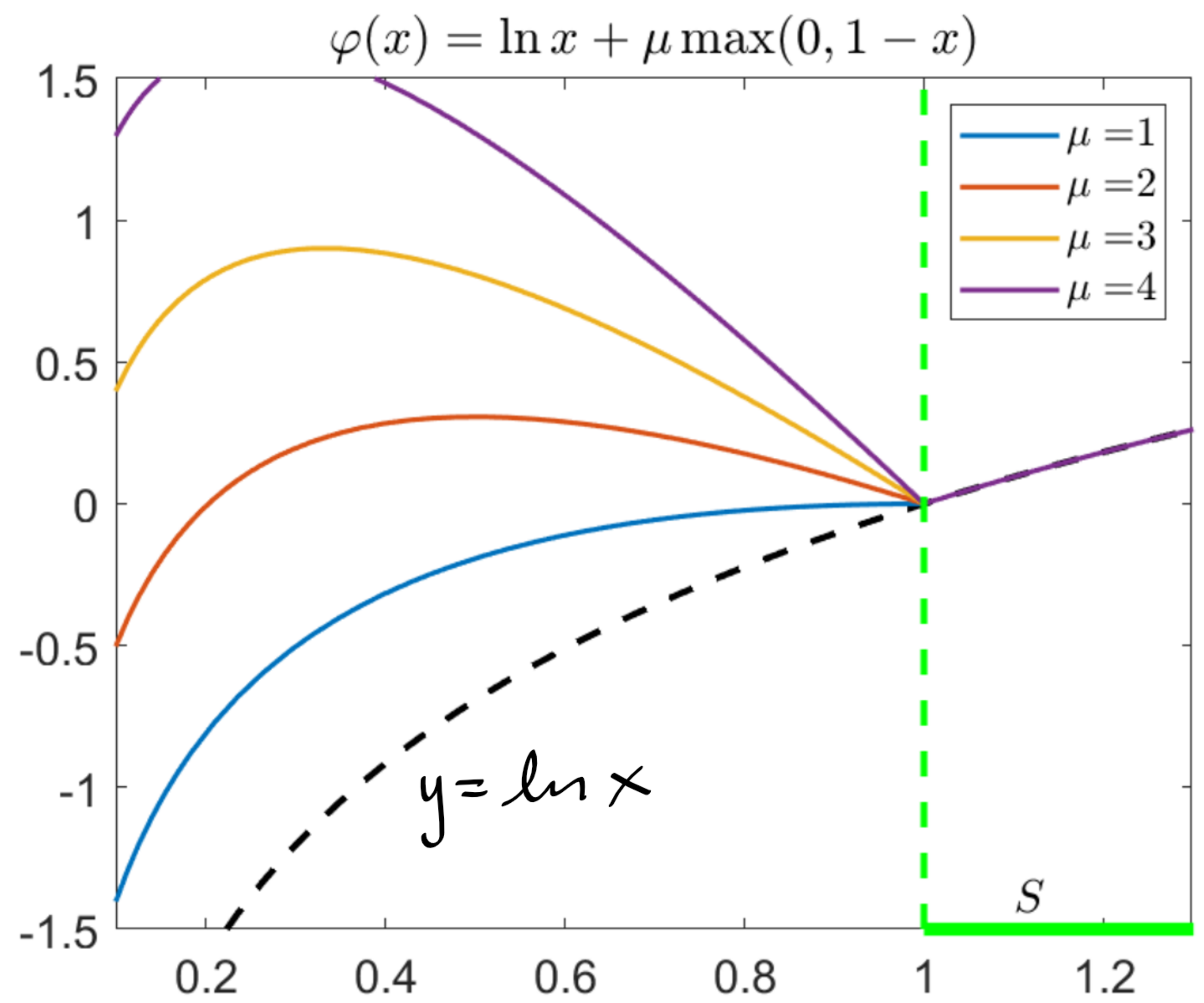
$g(x) = 1 - x \le 0$ $\qquad \mathbb{X} = \{x > 0\}$

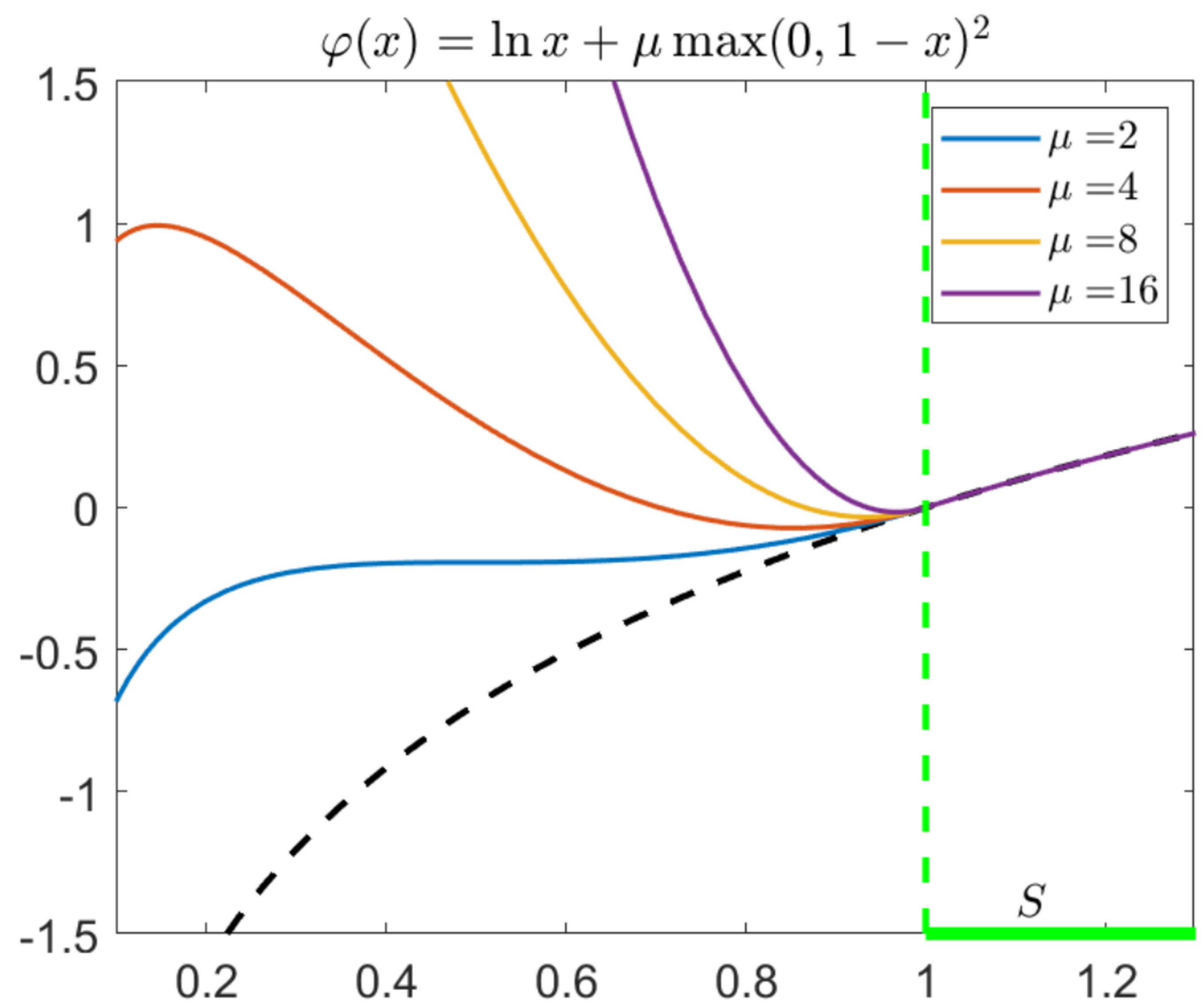
$$\varphi(x) = f(x) + \mu\,(\max(0, 1-x))^2$$

$$\varphi'(x) = \frac{1}{x} + 2\mu \max(0, 1-x)\,(-1) = \begin{cases} \frac{1}{x}, & x \ge 1 \\ \frac{1}{x} + 2\mu(x-1), & 0 < x < 1 \end{cases}$$

$$0 = \varphi'(x) = \frac{1}{x}\left(1 + 2\mu x(x-1)\right) = \frac{1}{x}\left(1 + 2\mu x^2 - 2\mu x\right)$$

$$= \frac{2\mu}{x}\left(\frac{1}{2\mu} + x^2 - x\right) \iff x = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{1}{2\mu}} \xrightarrow[\mu \to \infty]{} \frac{1}{2} \pm \frac{1}{2} = \begin{cases} 0 \\ 1 \end{cases}$$





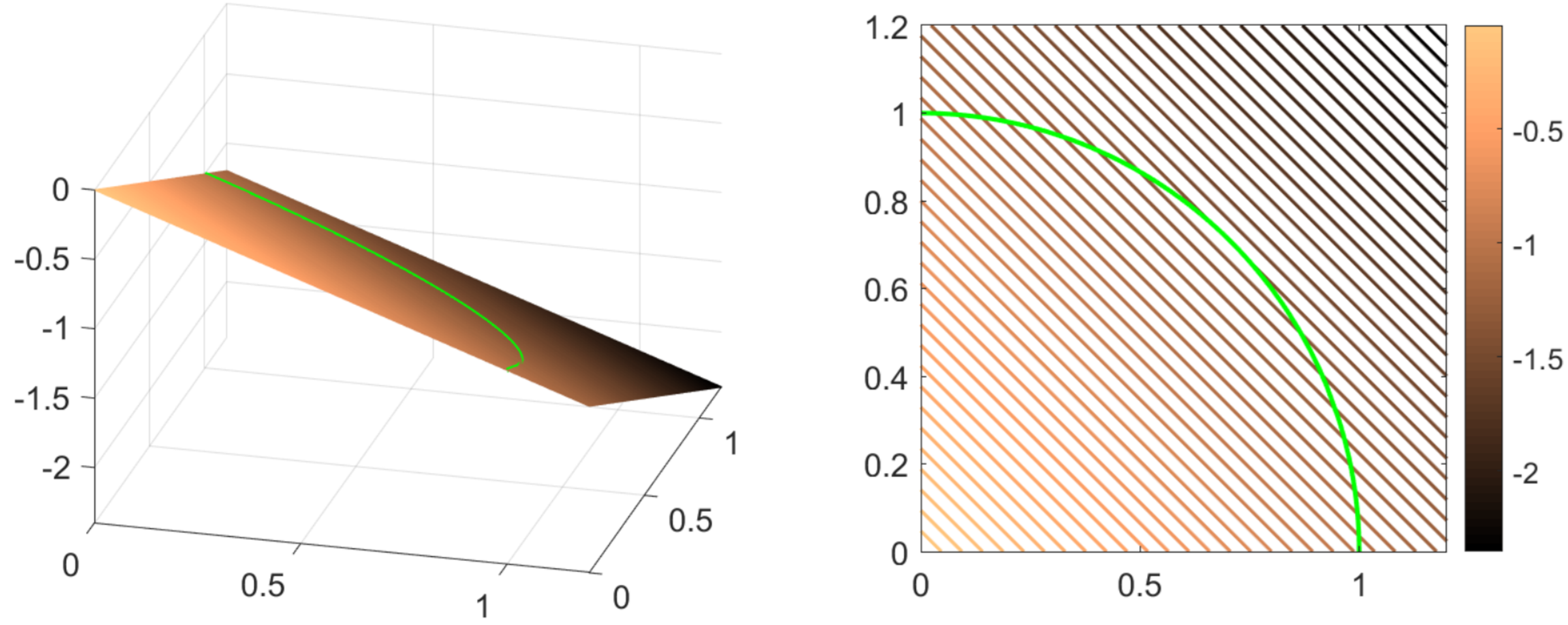
For $\mu \ge 2$ the minimizer $x < 1$ which is outside S

Ex. in 2D: $f(x) = -x_1 - x_2$, $h(x) = x_1^2 + x_2^2 - 1 = 0$

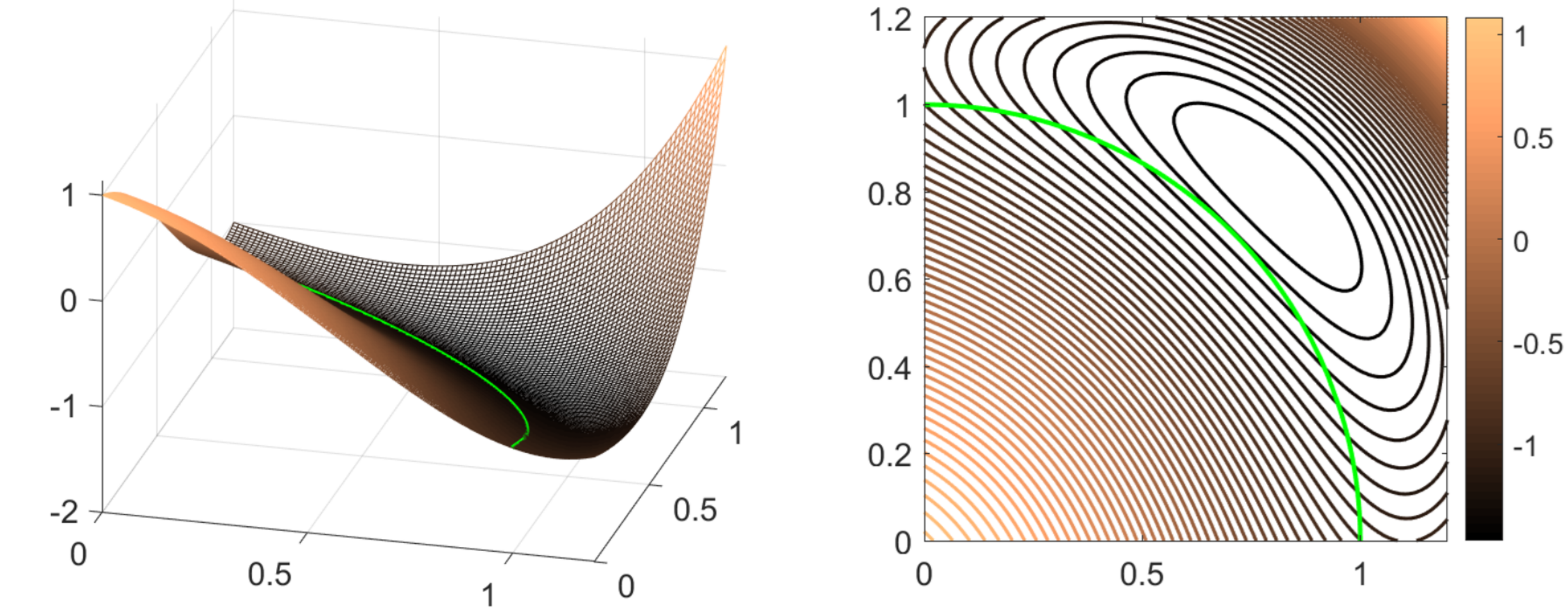
$\varphi(x)$

$\downarrow$

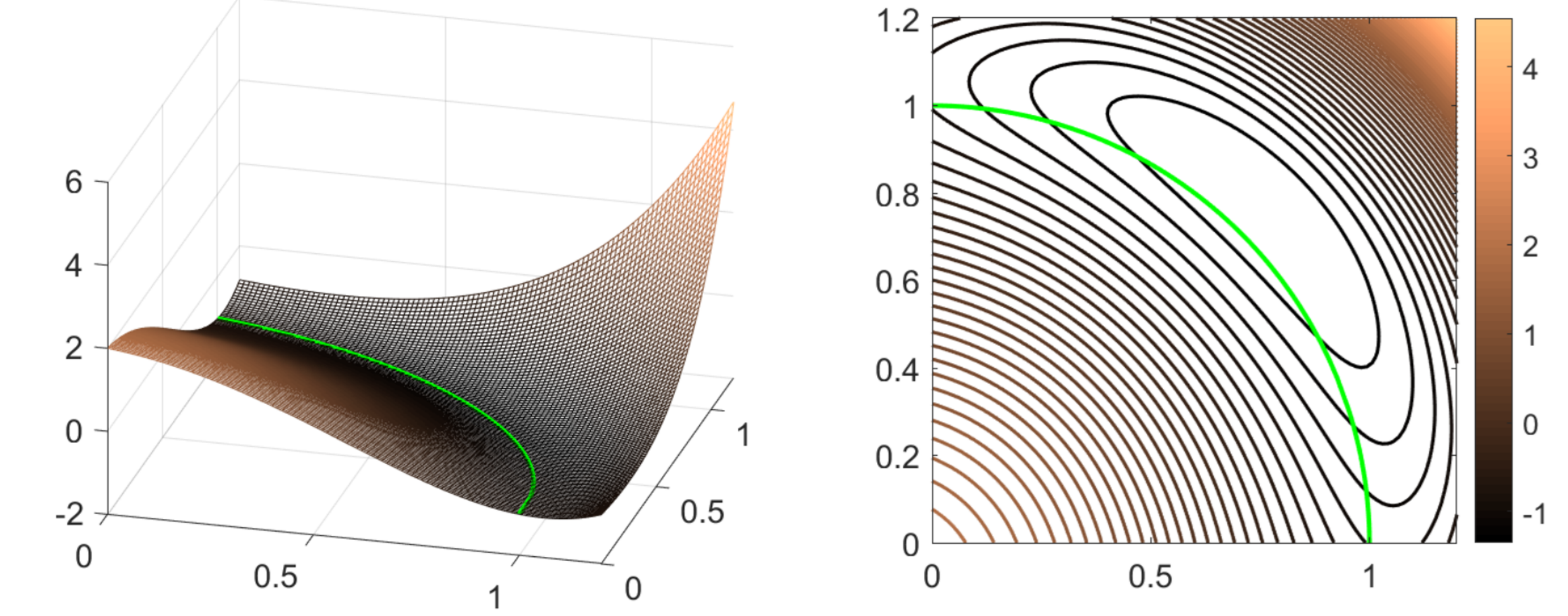
$\varphi(x_1, x_2) = -x_1 - x_2 + \mu(x_1^2 + x_2^2 - 1)^2$ with $\mu = 0$



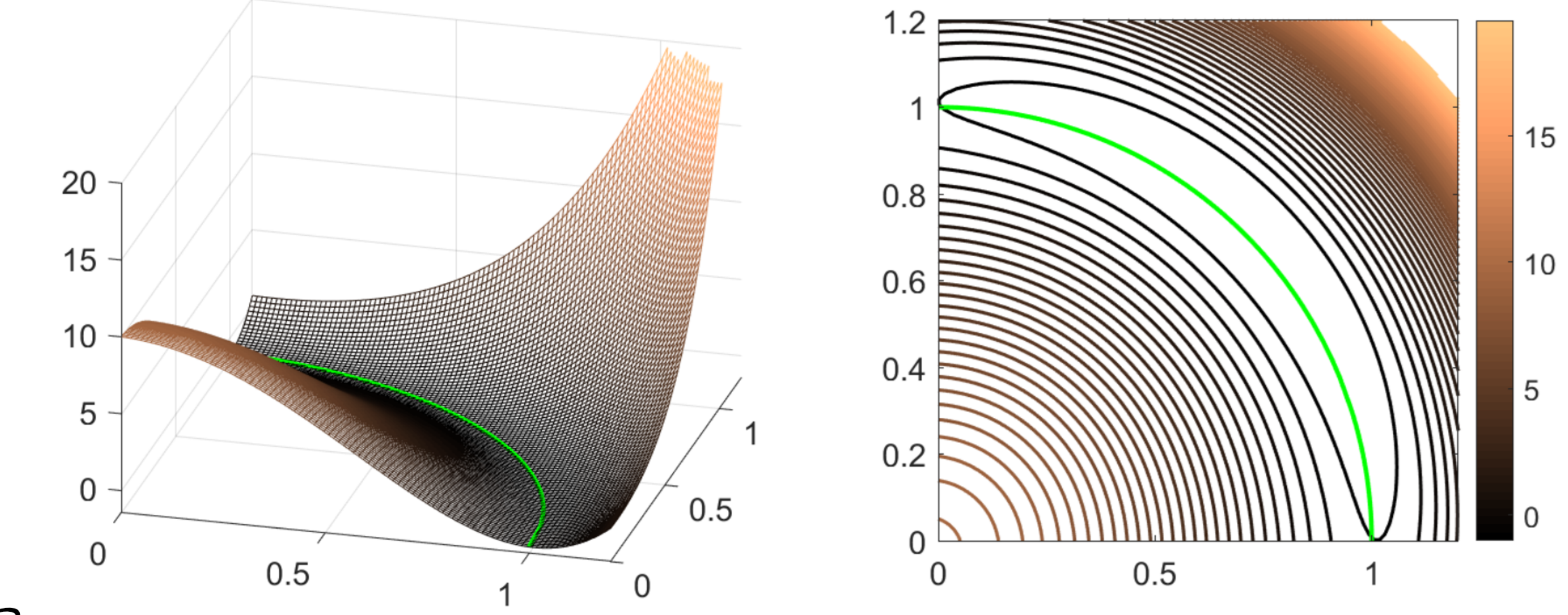
$\varphi(x_1, x_2) = -x_1 - x_2 + \mu(x_1^2 + x_2^2 - 1)^2$ with $\mu = 1$



$\varphi(x_1, x_2) = -x_1 - x_2 + \mu(x_1^2 + x_2^2 - 1)^2$ with $\mu = 2$



$\varphi(x_1, x_2) = -x_1 - x_2 + \mu(x_1^2 + x_2^2 - 1)^2$ with $\mu = 10$



Problem: long flat valley for large μ

$$\varphi(x) = -x_1 - x_2 + \mu\,(x_1^2 + x_2^2 - 1)^2$$

$$0 = \nabla\varphi(x) = \begin{pmatrix} -1 - 2\mu\,(x_1^2+x_2^2-1)\,2x_1 \\ \\ \end{pmatrix} \iff$$

$$\begin{cases} 4\mu x_1 (x_1^2+x_2^2-1) = 1 \\ 4\mu x_2 (x_1^2+x_2^2-1) = 1 \end{cases} \iff \begin{cases} 4\mu x_1 (x_1^2+x_2^2-1) = 1 \\ x_1 = x_2 \end{cases}$$

$\Rightarrow \quad x_1(2x_1^2-1) = \frac{1}{4\mu} \to 0$ as $\mu \to \infty$ and then

$$x_1 = x_2 = \frac{1}{\sqrt{2}}$$

Method: Choose an increasing sequence $\{\mu_k\}_{k=1}^{\infty}$ eg. $1, 10, 10^2, \dots$

- for each μ_k, $\underset{x \in X}{\text{minimize}}\ \varphi(x;\mu_k)$ with unconstrained optimization methods to get x_k.
- use x_k as starting point for next iteration.

<u>Thm 1</u>: If x_k exists for every k, i.e. $\varphi(x_k;\mu_k) \leq \varphi(x;\mu_k) \quad \forall x \in X$

and if (a subsequence of) $x_k \to \bar{x} \in X$ then $\bar{x}$ solves (P).

Consequences:

- If $\bar{x} \in \text{int}(S)$ (no $\overbrace{h_i(x) = 0}^{\text{constraints}}$ exist in (P)): since $x \in \text{int}(S)$ then $f(x) = \varphi(x;\mu)\ \forall\mu$, then $\bar{x}$ is found for μ_1.

- $\bar{x} \in \partial S$: then

$$f(x_k) \underset{\substack{\mu_k>0 \\ \alpha \geq 0}}{\leq} f(x_k) + \mu_k \alpha(x_k) = \varphi(x_k; \mu_k) \underset{x_k \text{ minimizer}}{\leq} \varphi(\bar{x}; \mu_k) \underset{\alpha(\bar{x})=0}{=} f(\bar{x})$$

<u>Either</u> $f(x_k) = f(\bar{x}) \Leftrightarrow \alpha(x_k) = 0 \Leftrightarrow$ <u>x_k solves (P)</u>

<u>or</u> $f(x_k) < f(\bar{x}) \Leftrightarrow \alpha(x_k) > 0 \Leftrightarrow$ <u>$x_k \notin S$</u>

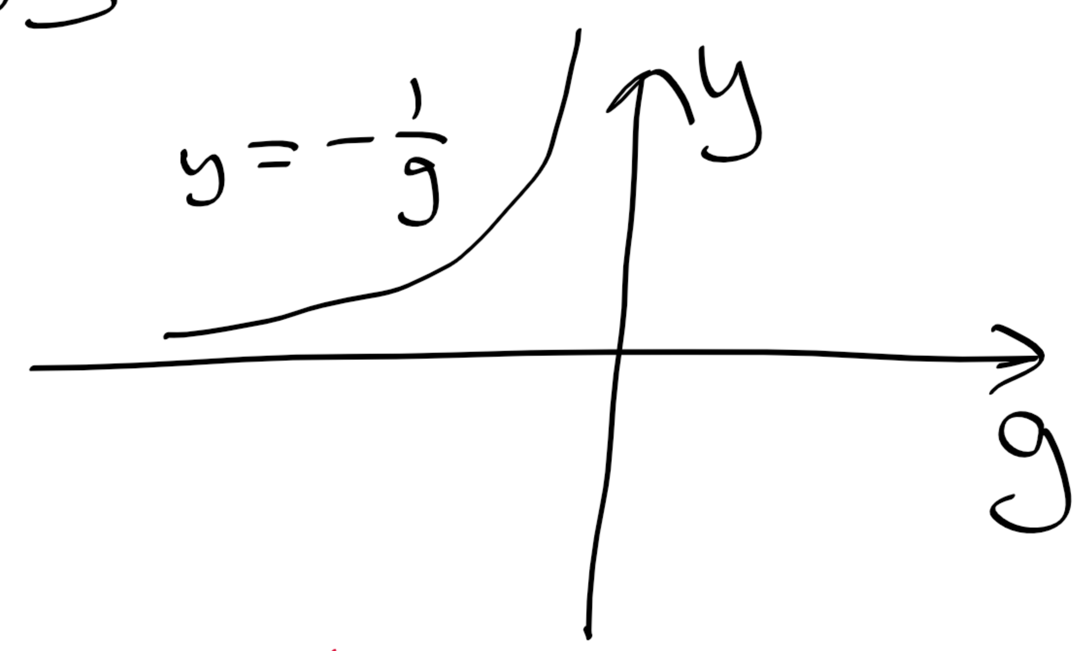
9.3 The barrier function method

(P) minimize $f(x)$
subject to $x \in S = \{x \in X : g(x) \leq 0\}$

The method is used when $x_k \notin S$ is not possible.

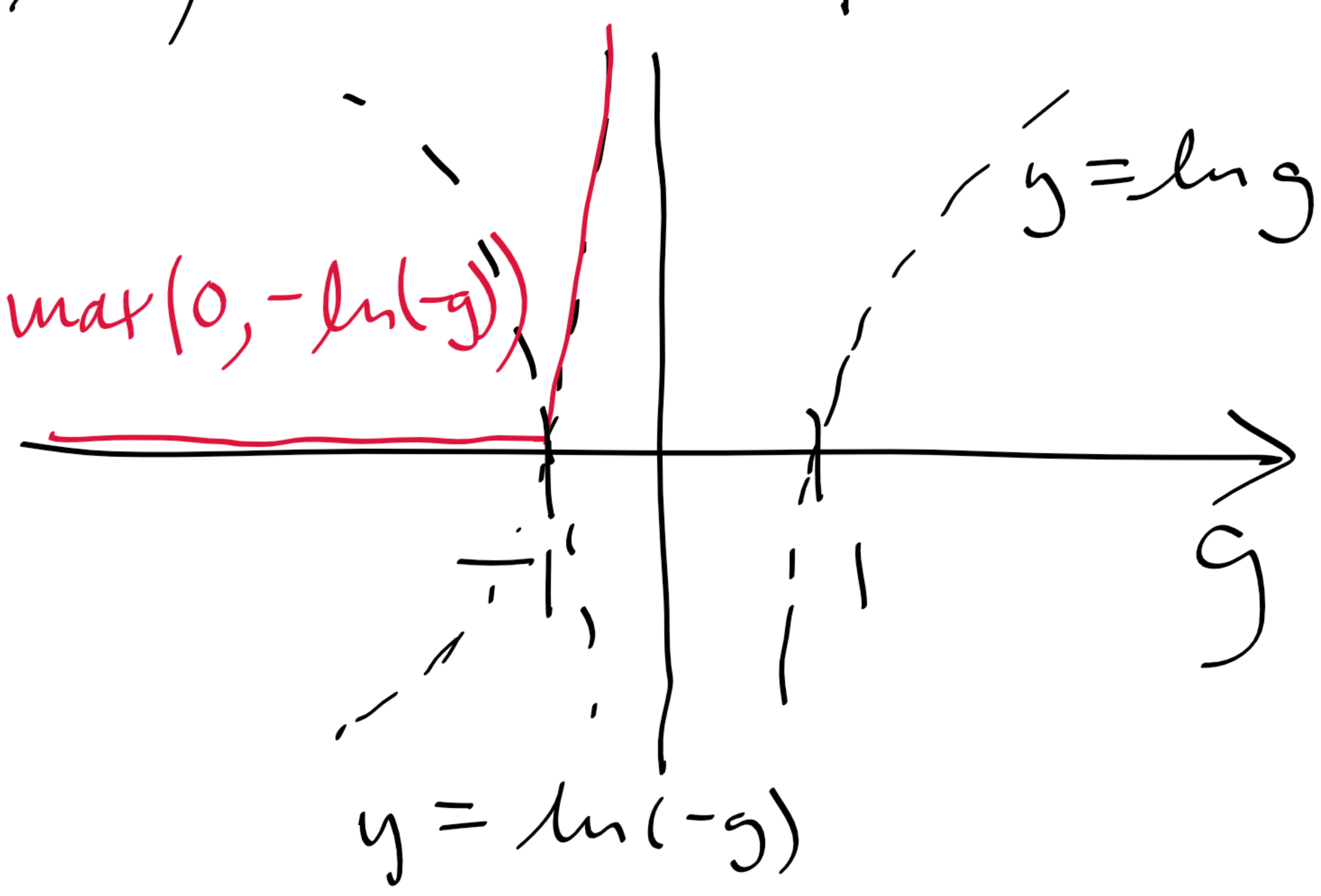
Form $\varphi(x) = f(x) + \varepsilon \underbrace{\beta(x)}_{\text{barrier function}}$, $\varepsilon > 0$

$\beta(x) \to \infty$ as $\text{int}(S) \ni x \to \partial S$



For ex. $\beta(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)}$

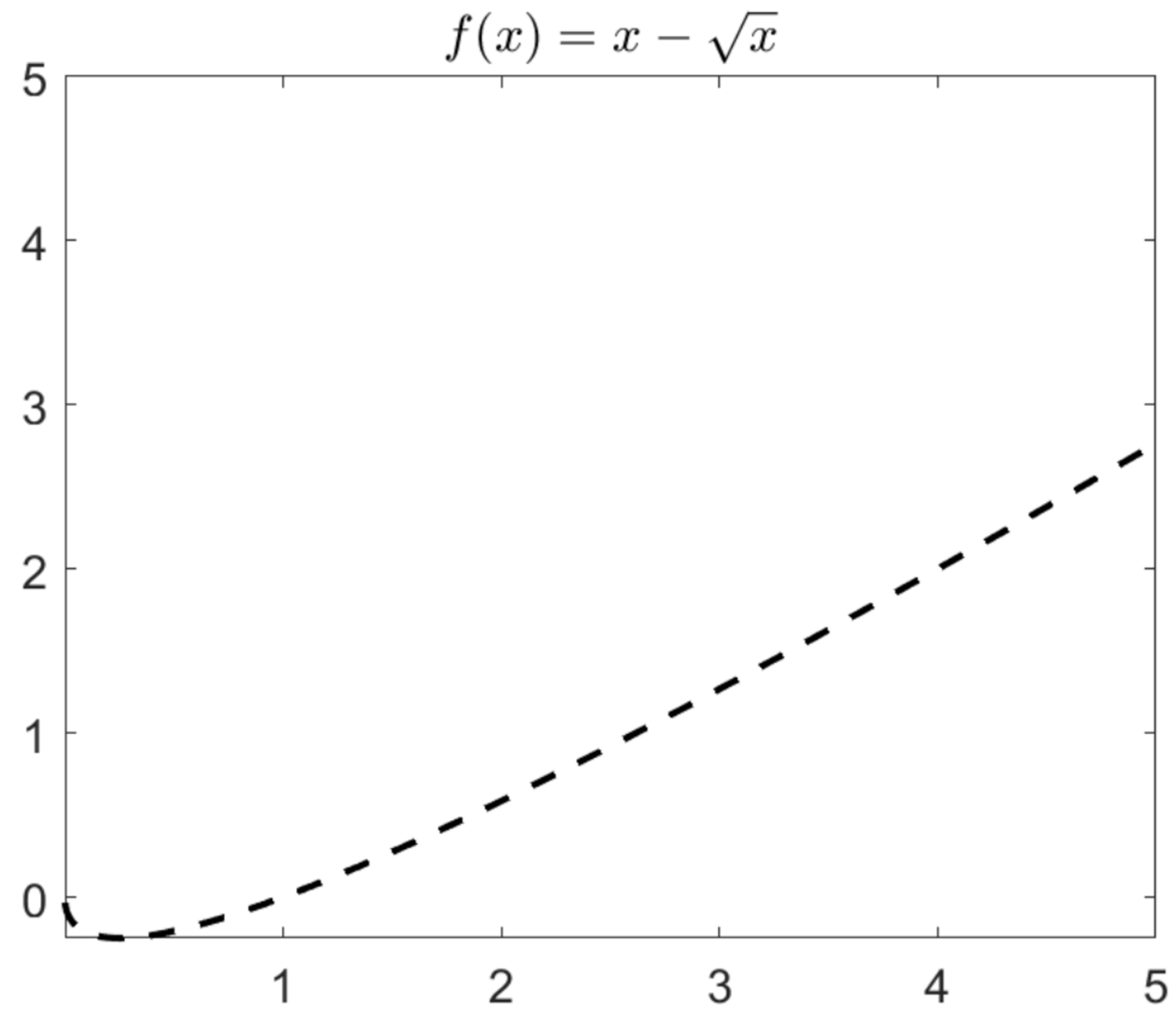
$\beta(x) = \sum_{i=1}^{m} \max(0, -\ln(-g_i(x)))$

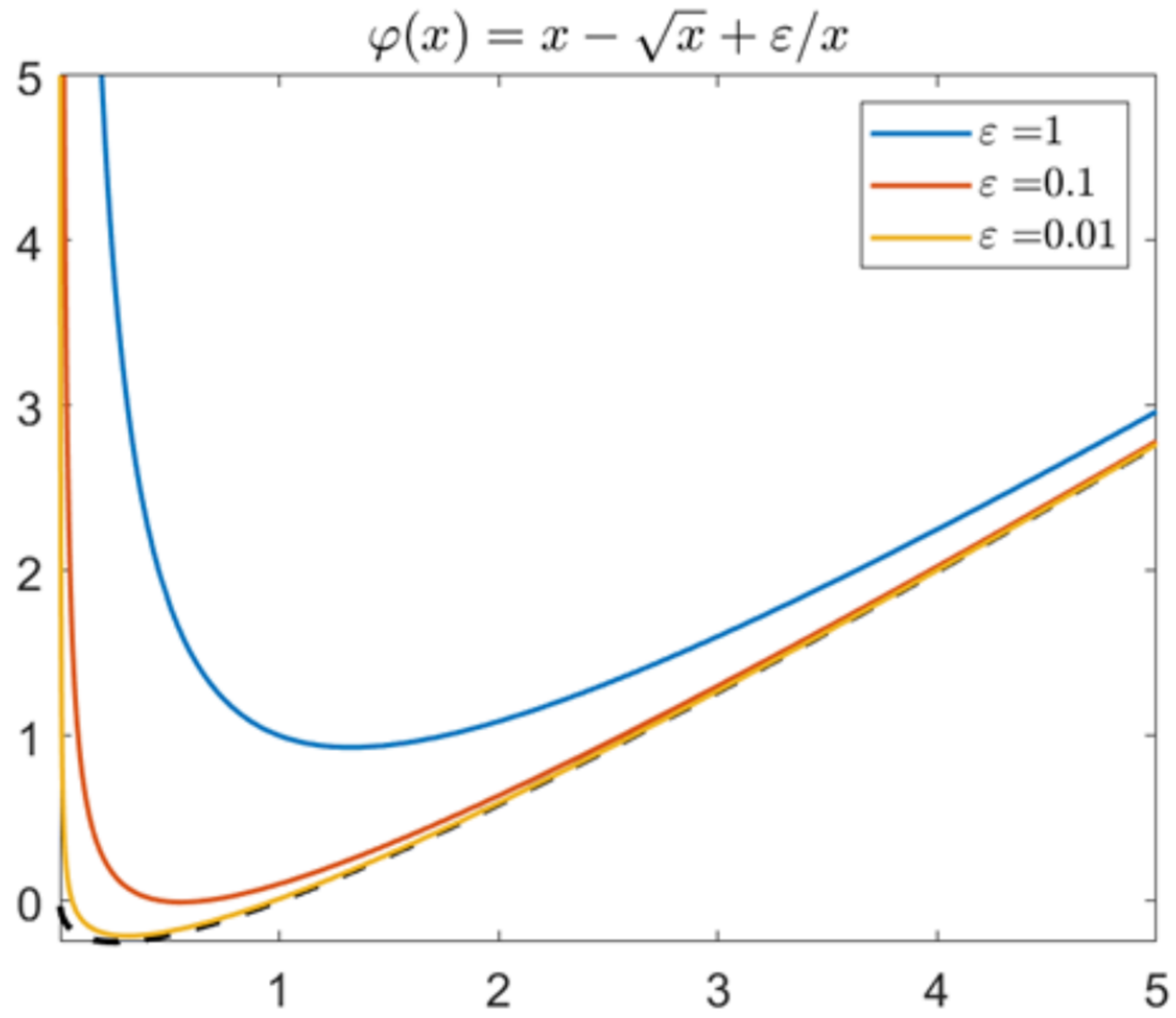


Method: Choose a decreasing sequence $\{\varepsilon_k\}$ etc.

Ex. in 1D: $f(x) = x - \sqrt{x}$
$g(x) = -x \leq 0$

$\varphi(x) = x - \sqrt{x} + \frac{\varepsilon}{x}$ $x > 0$





To avoid $x_k \notin S$, set $\beta(x) = \infty$, $x \notin S$.