Heuristics In Full-Waveform Inversion

Rafael Lago, Felix Herrmann

Consortium Meeting, Amsterdam, June 2014

Frequency Domain Full Waveform Inversion Overview

Full Waveform Inversion - Overview

Define the **misfit function** as

$$\min_{m \in \mathcal{A}} \Phi(m) = \sum_{i}^{n_s} \left\| d^i - P_r u^i \right\|_W .$$

for some norm W (usually L^2 with some regularization), and the **gradient** of the reduced formulation

$$\nabla \Phi(m) = \sum_{i}^{n_s} \mathcal{R} \left\{ (u^i)^T \left[\frac{\partial A}{\partial m} \right]^T w^i \right\} .$$

- d^i observed data
- u^i (approximated) computed data
- m Earth parameters; what we are trying to invert!

Forward Modelling - Overview

Forward problem:

$$u^i = A^{-1}(m)q^i$$

(computation of $\Phi(m)$)

Backward problem:

$$w^i = A^{-H}(m)P_r^H(d^i - u^i).$$

(computation of $\nabla\Phi(m)$)

Both require a PDE solve, computed with an (sufficiently large) accuracy ε.

- $A(m)$ operator governing the physics of Earth
- d^i, u^i observed and computed data
- P_r restricts the computed data to the receivers
- q^i source

Frugal FWI Overview

Gradient-Descent with Errors

Let

$$\nabla\tilde{\Phi}(m_k) = \nabla\Phi(m_k) + \boldsymbol{e_k}$$

for some **error** e_k. Then, for **strongly convex** problems:

$$\Phi(m_k) - \Phi(m_*) < a_k(\Phi(m_0) - \Phi(m_*))$$

$$a_k = max\left\{c^k, ||\boldsymbol{e_k}||_{\boldsymbol{2}}^{\boldsymbol{2}}\right\}, \qquad 0 \leq c \leq 1$$

where c is the condition number of the problem.

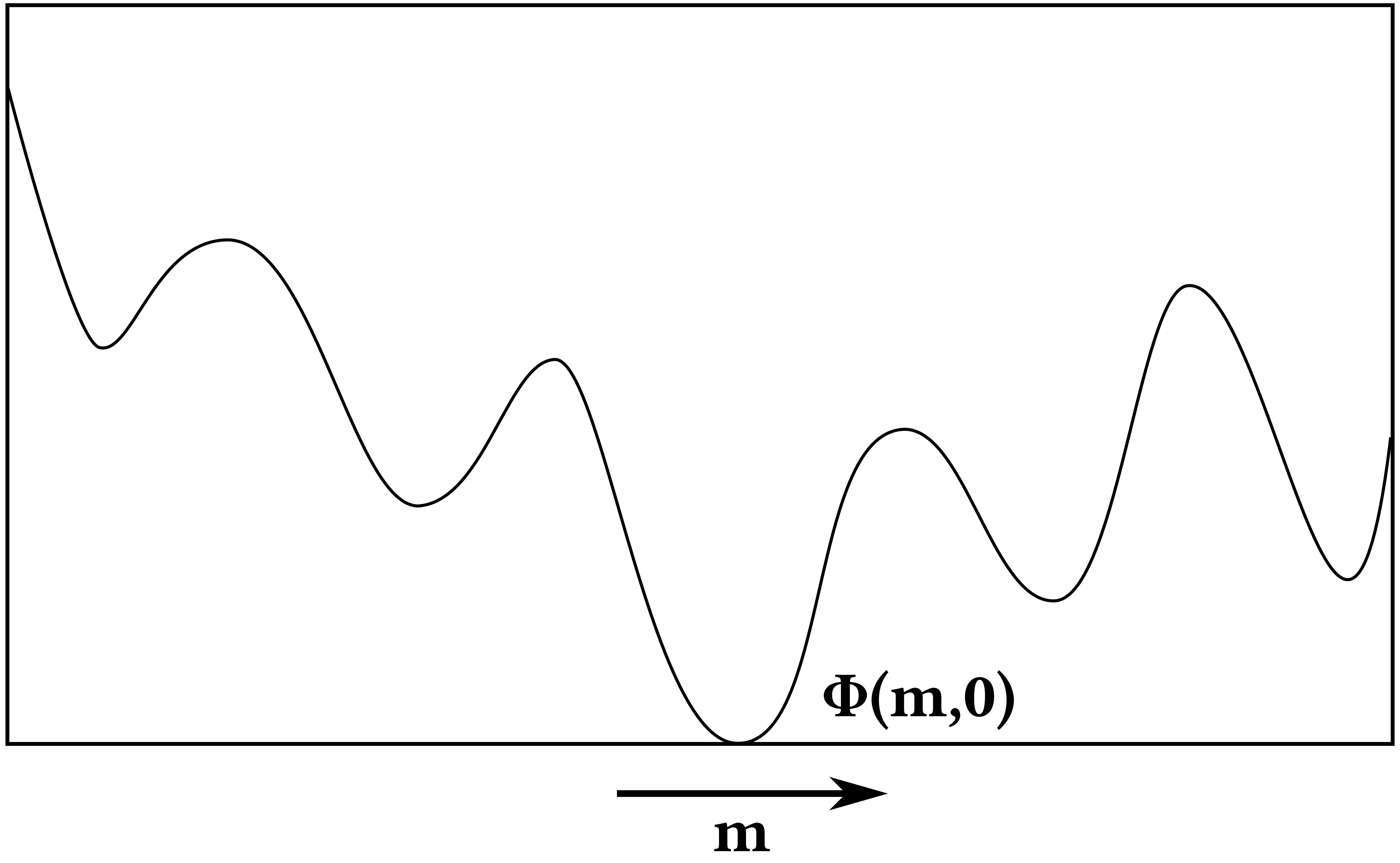
[Friedlander and Schmidt, 2012]

Relaxing the Physics - Approximating u^i and w^i

$$\Phi(m) = \sum_{i}^{n_s} \left\| d^i - \boldsymbol{u^i} \right\|_W$$

$$\boldsymbol{u^i} \approx A^{-1}(m) q^i$$

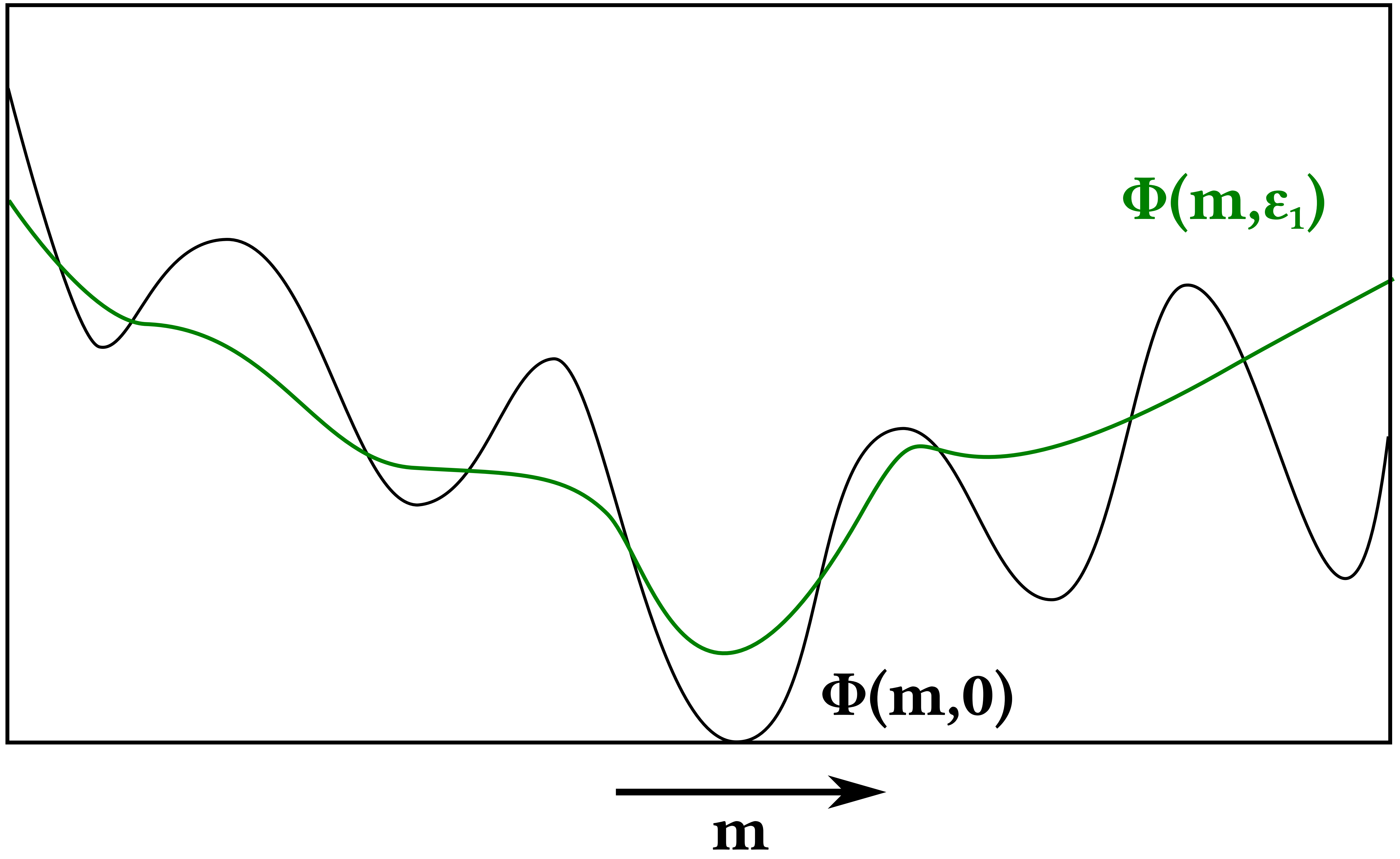
$$\frac{\left\| A(m)\boldsymbol{u^i} - q^i \right\|_2}{\left\| q^i \right\|_2} < \varepsilon$$



$$\Phi(m) = \sum_{i}^{n_s} \left\| d^i - \boldsymbol{u^i} \right\|_W$$

$$\boldsymbol{u^i} \approx A^{-1}(m) q^i$$

$$\frac{\left\| A(m)\boldsymbol{u^i} - q^i \right\|_2}{\left\| q^i \right\|_2} < \varepsilon$$



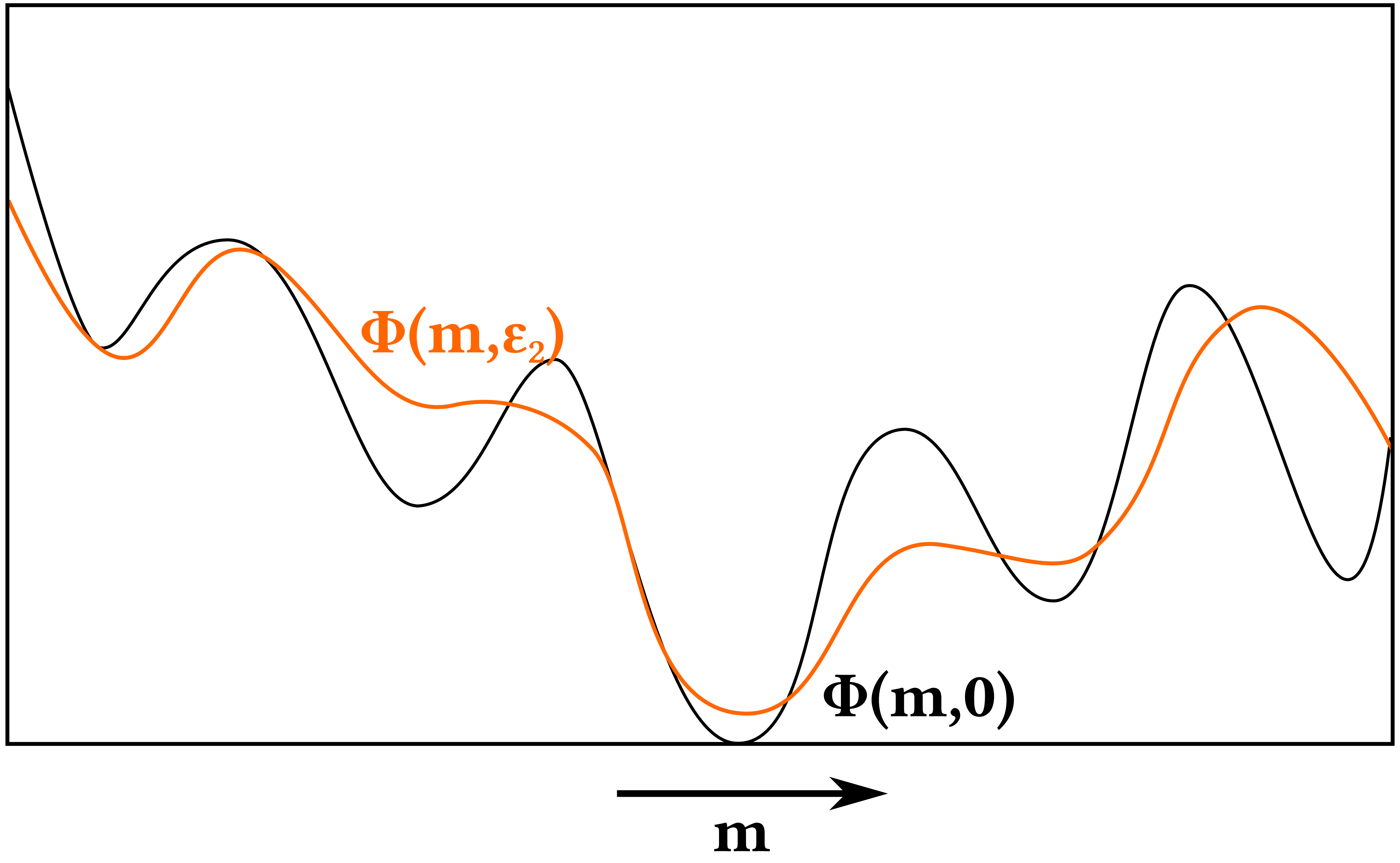
We hope that "large" ε_k can **"convexify"** Φ

Relaxing the Physics - Approximating u^i and w^i

$$\Phi(m) = \sum_{i}^{n_s} \left\| d^i - \boldsymbol{u^i} \right\|_W$$

$$\boldsymbol{u^i} \approx A^{-1}(m) q^i$$

$$\frac{\left\| A(m)\boldsymbol{u^i} - q^i \right\|_2}{\left\| q^i \right\|_2} < \varepsilon$$



We hope that "large" ε_k can **"convexify"** Φ
and that $\Phi(m, \varepsilon_k)$ **"converges"** to $\Phi(m, 0)$

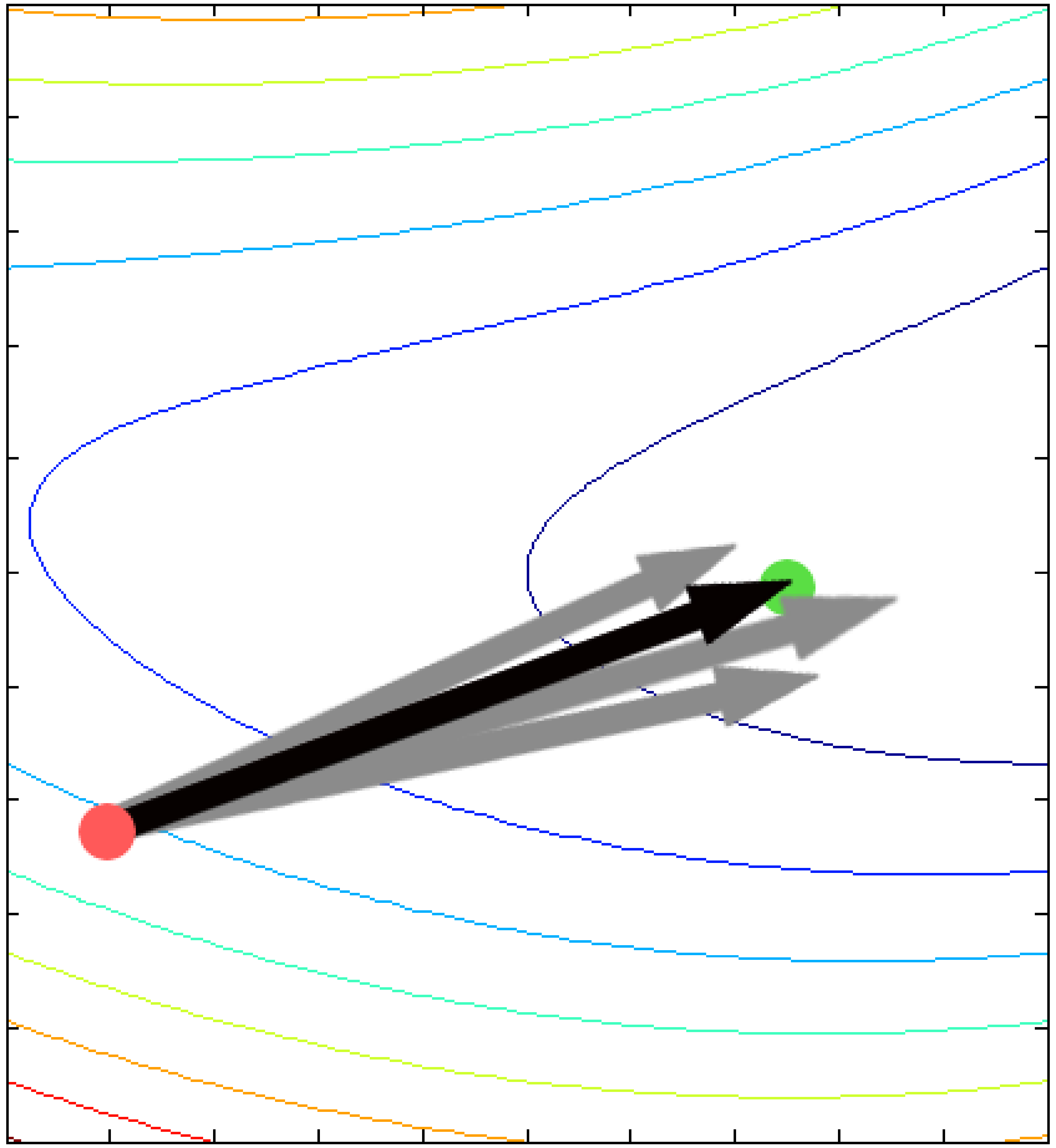
Choosing ε_k - Approximating u^i and w^i

$$|\Phi(m, \alpha^k \epsilon) - \Phi(m, \boldsymbol{\alpha}^{\boldsymbol{k+1}} \boldsymbol{\epsilon})| \leq \boldsymbol{\eta} \Phi(m, \boldsymbol{\alpha}^{\boldsymbol{k+1}} \boldsymbol{\epsilon})$$

$$\boldsymbol{u}^{\boldsymbol{i}} \approx A^{-1}(m) q^i$$

$$\frac{\left\| A(m) \boldsymbol{u}^{\boldsymbol{i}} - q^i \right\|_2}{\left\| q^i \right\|_2} < \boldsymbol{\alpha}^{\boldsymbol{k+1}} \boldsymbol{\epsilon}$$

(use the final tolerance to compute w^i)



Chosen Parameters

- $\alpha = 0.5$
- $\epsilon = 10^{-2}$
- $\eta = 5 \times 10^{-2}$

[Herrmann et al., 2013]

Frugal FWI - Source Subsampling

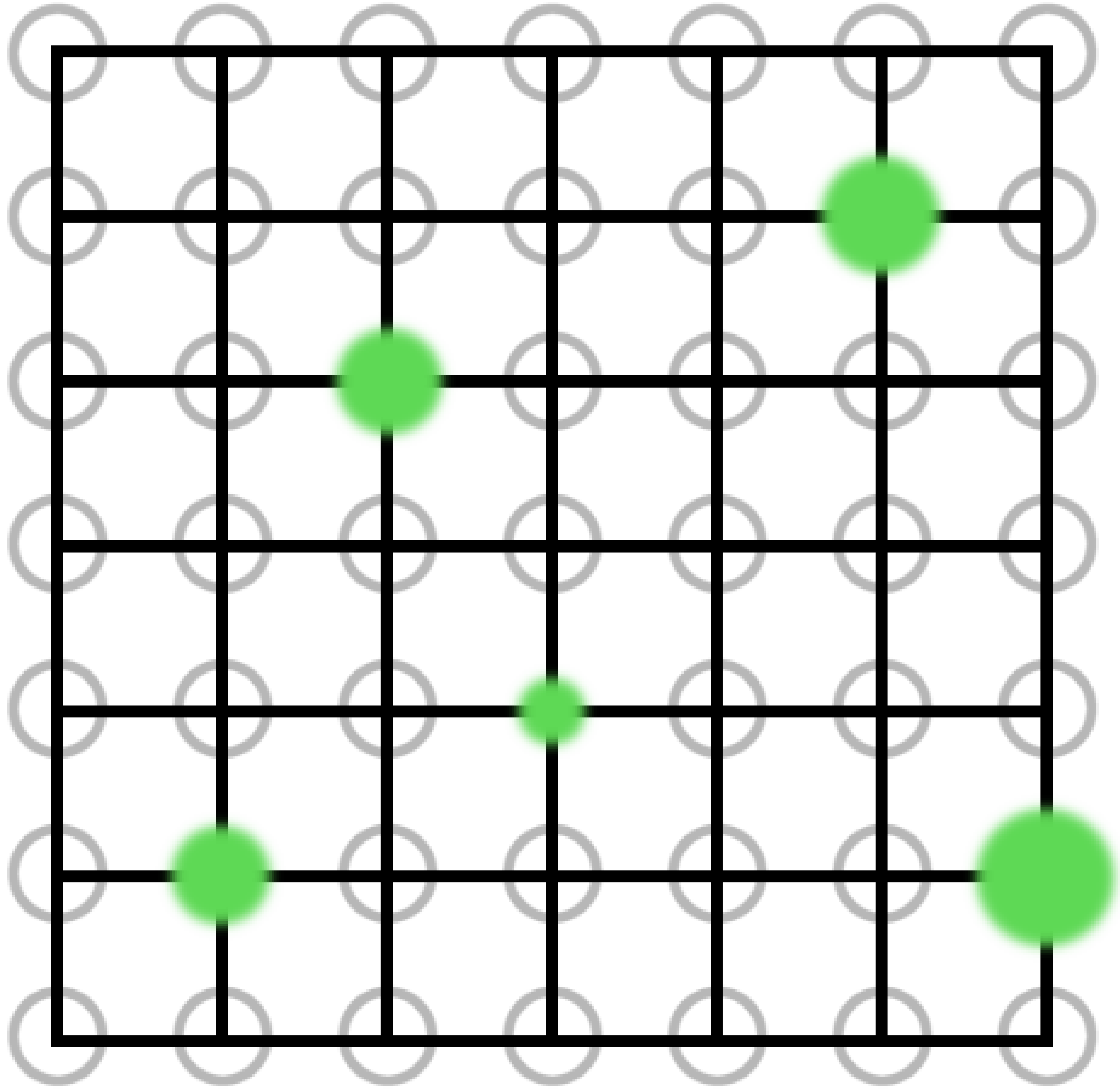
$$\tilde{\Phi}(m) = \sum_{i \in \mathcal{I}_k}^{b_k} \left\| d^i - u^i \right\|_W$$

$$\mathcal{I}_k \subset \{1, 2, ..., n_s\}, \mathcal{I}_k^{\#} = b_k$$

$\mathcal{I}_k$ is chosen at random *without replacement*. The expected error is given by

$$\|e_k\|_2 \propto \sqrt{\frac{1}{b_k} - \frac{1}{n_s}}$$

$$b_k \sim \min\left\{ \left(\epsilon^k + \frac{1}{n_s} \right)^{-1}, n_s \right\}.$$



[Friedlander and Schmidt, 2012] and [Herrmann et al., 2013]

CGMN, CRMN & Kaczmarz Sweep

Kaczmarz (Double) Sweep - Overview

$$u_{i+1} = u_i + \frac{\gamma(q_i - a_i^H u_i)a_i}{\|a_i\|_2^2}$$

- q_i — ith element of q
- a_i — ith row of A as a column vector
- γ — relaxation parameter $\in (0, 2)$

Kaczmarz sweeps **guarantees convergence** in a finite (possibly **large**) number of steps.

Kaczmarz (Double) Sweep - Overview

$$Q = Q_1 Q_2 ... Q_N Q_N Q_{N-1} ... Q_1 \qquad Q_i = I - \frac{\gamma}{\|a_i\|_2^2} a_i a_i^H$$

$$u := Qu + Rq \qquad \Longrightarrow \qquad (I - Q)u = Rq$$

- Q is **symmetric positive definite**
- We can use **CG** to solve this system
- Neither Q nor R need to be computed in practice
- Equivalent to using **CG on the normal equations**, preconditioned by SSOR

1 (CGMN).

$p_0 = r_0 = dkswp(A, u_0, b, \gamma) - u_0;$

while *not converged* **do**

  $q_k = p_k - dkswp(A, p_k, 0, \gamma);$

  $\alpha_k = \langle r_k, r_k \rangle \, / \, \langle p_k, q_k \rangle;$

  $u_{k+1} = u_k + \alpha_k r_k;$

  $r_{k+1} = r_k - \alpha_k q_k;$

  $\beta_k = \langle r_{k+1}, r_{k+1} \rangle \, / \, \langle r_k, r_k \rangle;$

  $p_{k+1} = r_k + \beta_k p_k;$

  $k = k + 1;$

end while

- Very low memory cost
- Very simple implementation
- Suitable for any matrix A (even nonsquare)

On CG:[Hestenes and Stiefel, 1952], on CGMN: [Björck and Elfving, 1979]

2 (CRMN).

while *not converged* **do**

$Ar_k := r_k - dkswp(A, r_k, 0, \gamma)$;

$\beta_k = \langle r_k, Ar_k \rangle \,/\, \langle r_{k-1}, Ar_{k-1} \rangle$;

$p_k = r_k + \beta_k p_{k-1}$;

$Ap_k = Ar_k + \beta_k Ap_{k-1}$;

$\alpha_k = \langle r_k, Ar_k \rangle \,/\, \langle Ap_k, Ap_k \rangle$;

$u_{k+1} = u_k + \alpha_k r_k$;

$r_{k+1} = r_k - \alpha_k q_k$;

$k = k + 1$;

end while

- Very **low memory cost**
- One **extra** vector storage
- One **extra** inner product
- **Minimal residual** properties

On CR: [Stiefel, 1955], comparisons:[Fong and Saunders, 2012, Eiermann and Ernst, 2001]

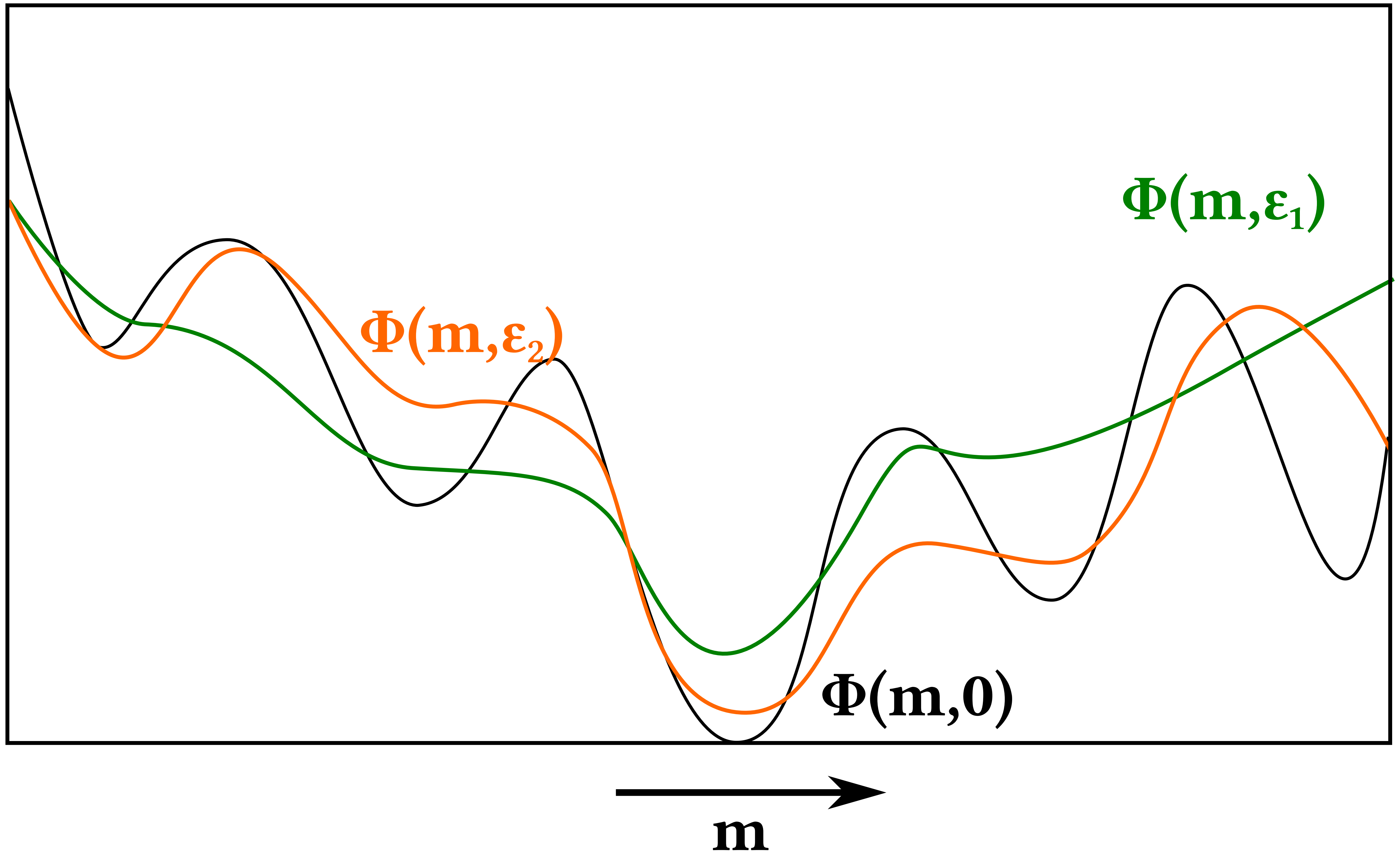
On Heuristics

A Matter of Robustness

$$\frac{|\Phi(m,\alpha^k\epsilon)-\Phi(m,\boldsymbol{\alpha^{k+1}\epsilon})|}{\Phi(m,\boldsymbol{\alpha^{k+1}\epsilon})}\leq\boldsymbol{\eta}$$

$$\boldsymbol{u^i}\approx P_rA^{-1}(m)q^i$$

$$\frac{\left\|A(m)\boldsymbol{u^i}-q^i\right\|_2}{\left\|q^i\right\|_2}<\boldsymbol{\alpha^{k+1}\epsilon}$$



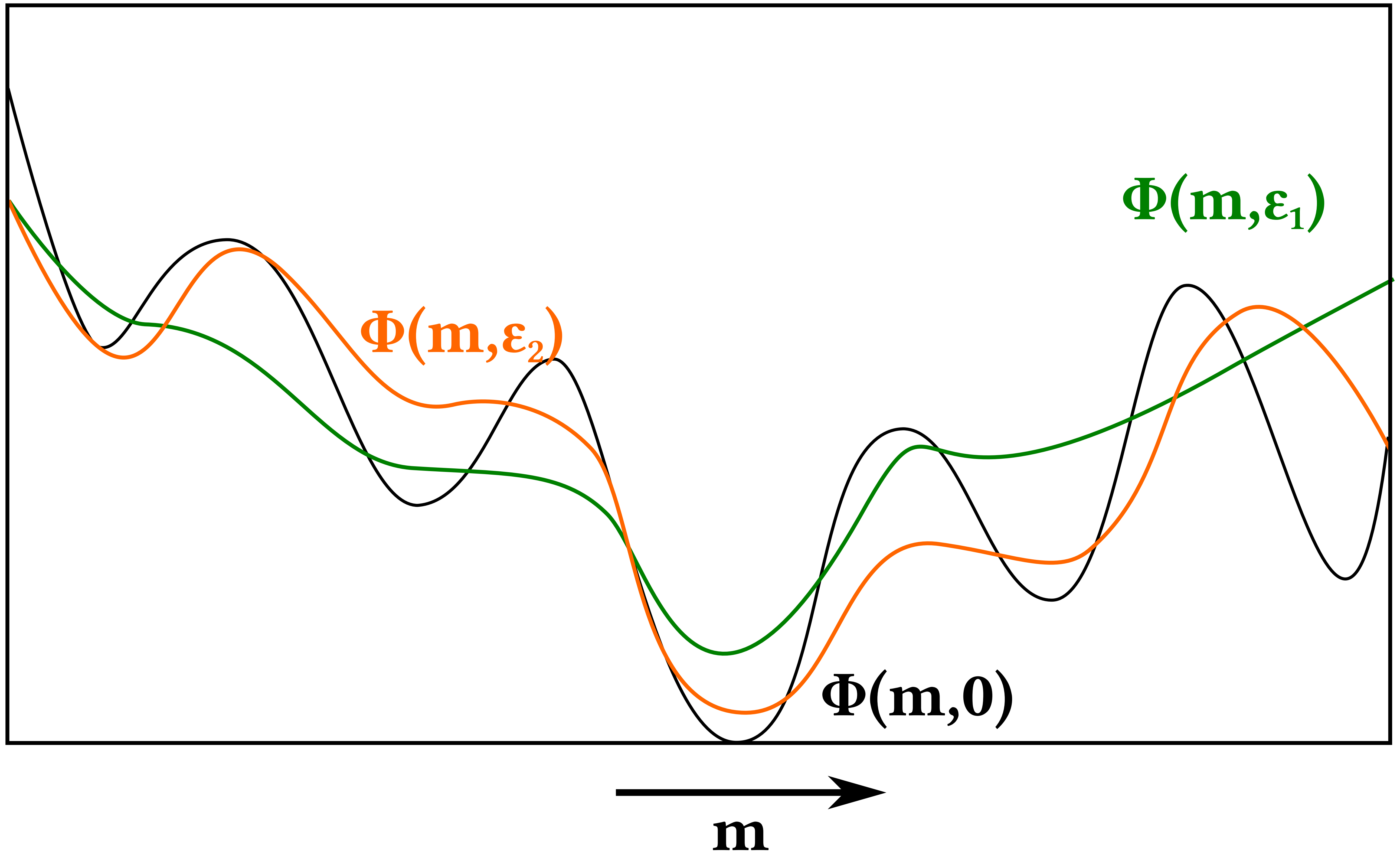
We hope that "large" ε_k can "convexify" Φ and that $\Phi(m,\varepsilon_k)$ "converges" to $\Phi(m,0)$

A Matter of Robustness

$$\frac{|\Phi(m,\alpha^k\epsilon) - \Phi(m,\boldsymbol{\alpha^{k+1}\epsilon})|}{\Phi(m,\boldsymbol{\alpha^{k+1}\epsilon})} \leq \boldsymbol{\eta}$$

$$\boldsymbol{u^i} \approx P_r A^{-1}(m) q^i$$

$$\frac{\left\| A(m)\boldsymbol{u^i} - q^i \right\|_2}{\left\| q^i \right\|_2} < \boldsymbol{\alpha^{k+1}\epsilon}$$

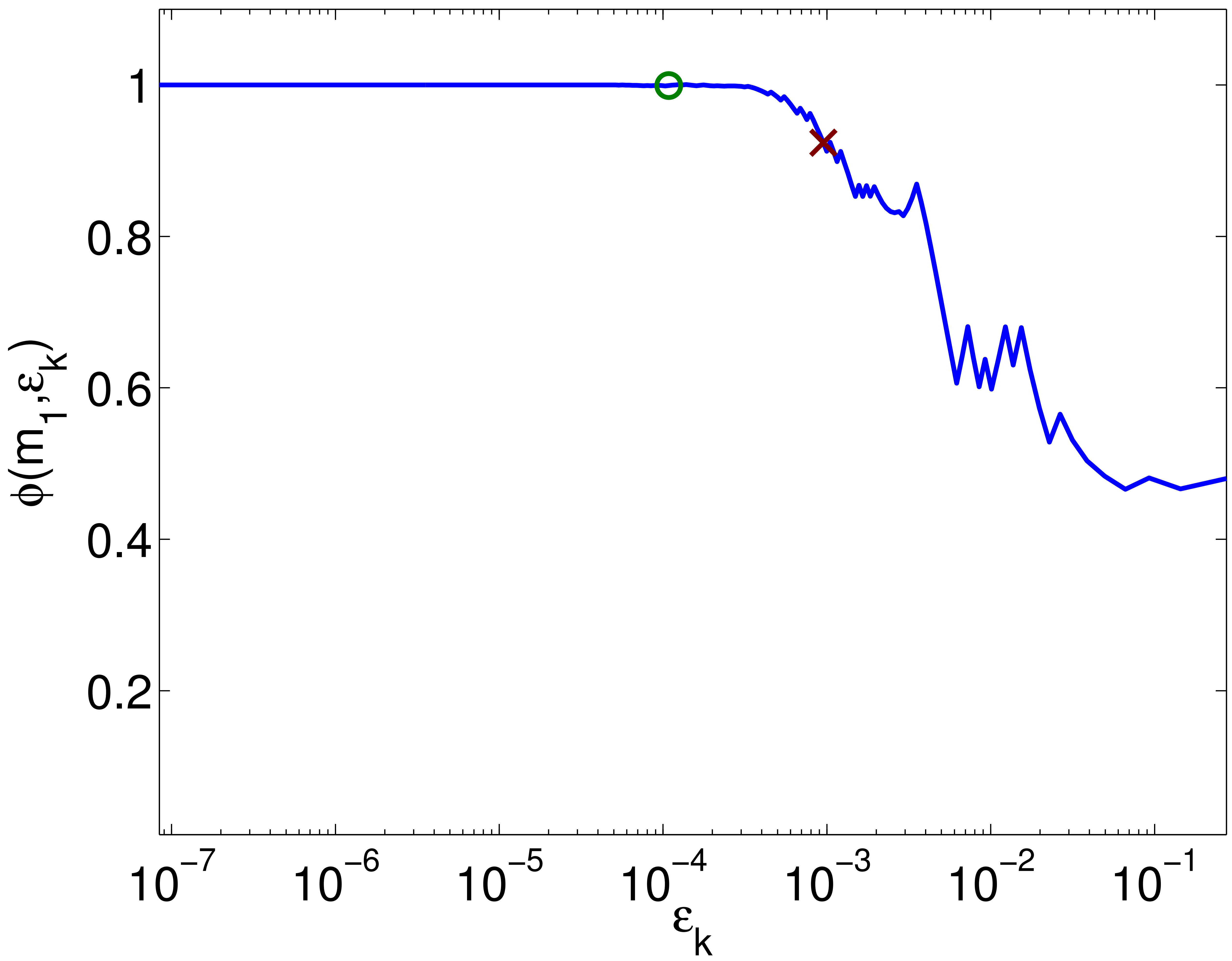


We hope that "large" ε_k can "convexify" Φ
and that $\Phi(m,\varepsilon_k)$ "converges" to $\Phi(m,0)$
But... does this happen? **How quickly? How consistently?**

A Glance Into The Misfit

o - relative residual at 10^{-4}
x - relative residual at 10^{-3}

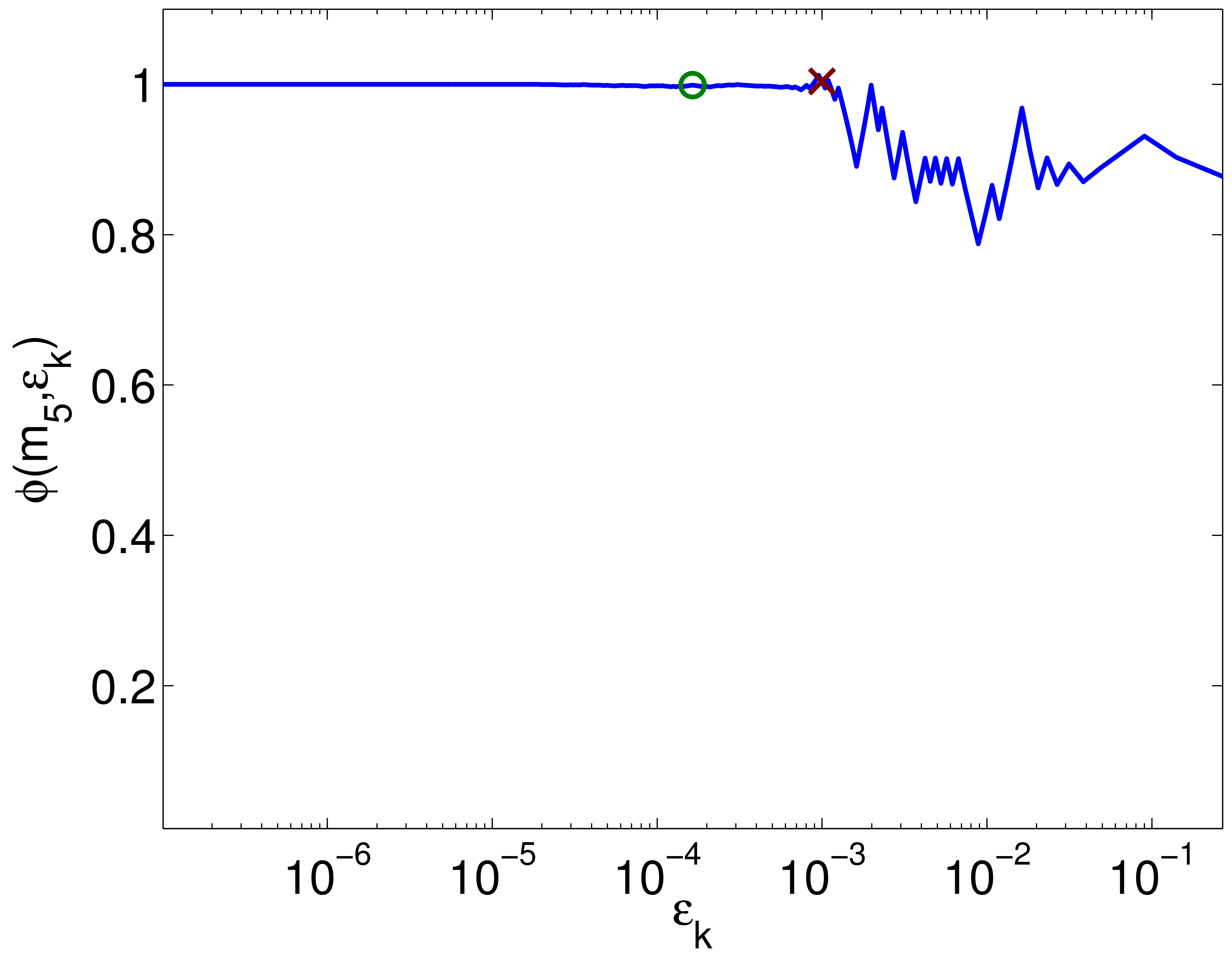
Misfit as a function of ε_k for the **1st** iteration of lBFGS



A Glance Into The Misfit

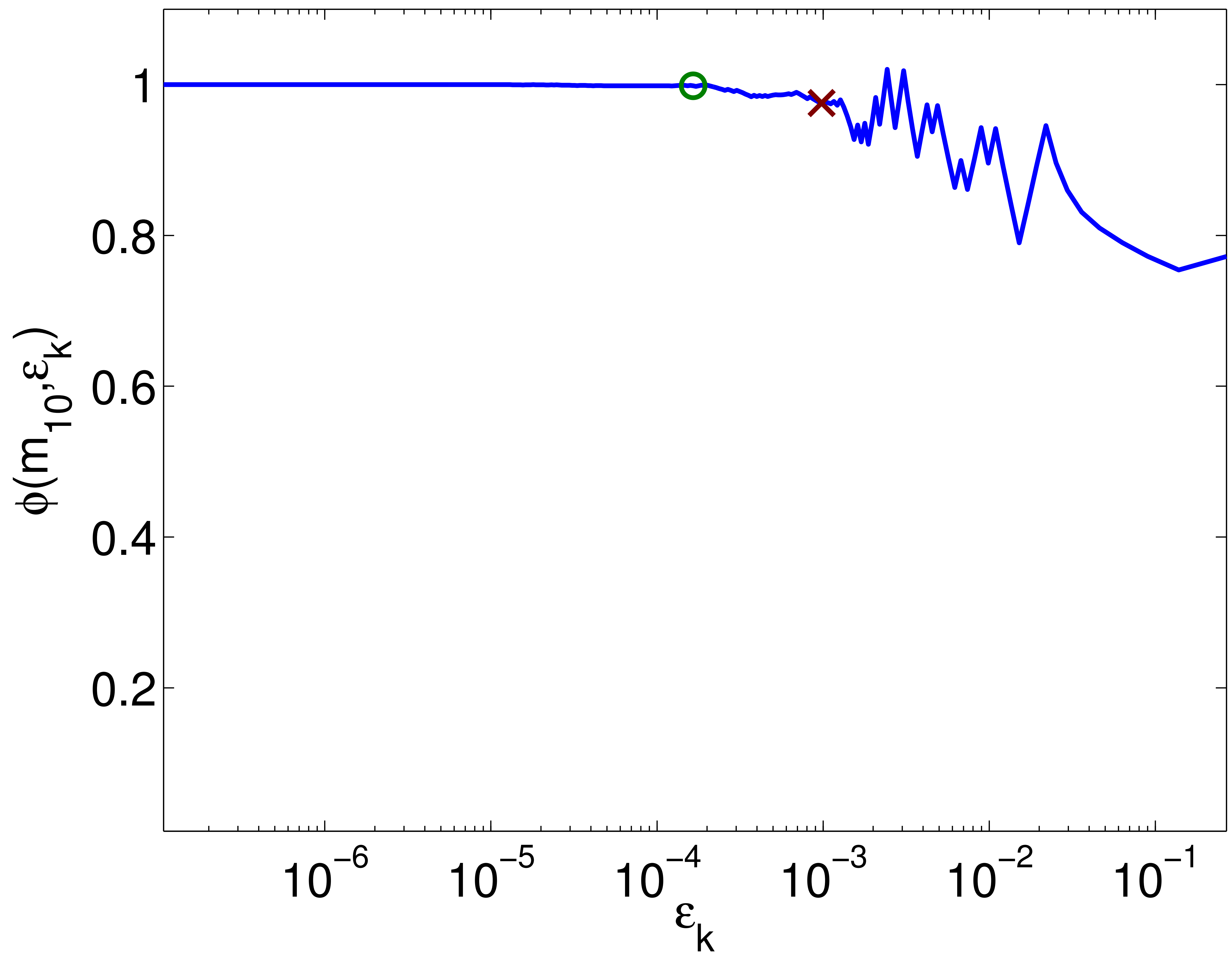
o - relative residual at 10^{-4}
x - relative residual at 10^{-3}

Misfit as a function of ε_k for the **5th** iteration of lBFGS



A Glance Into The Misfit

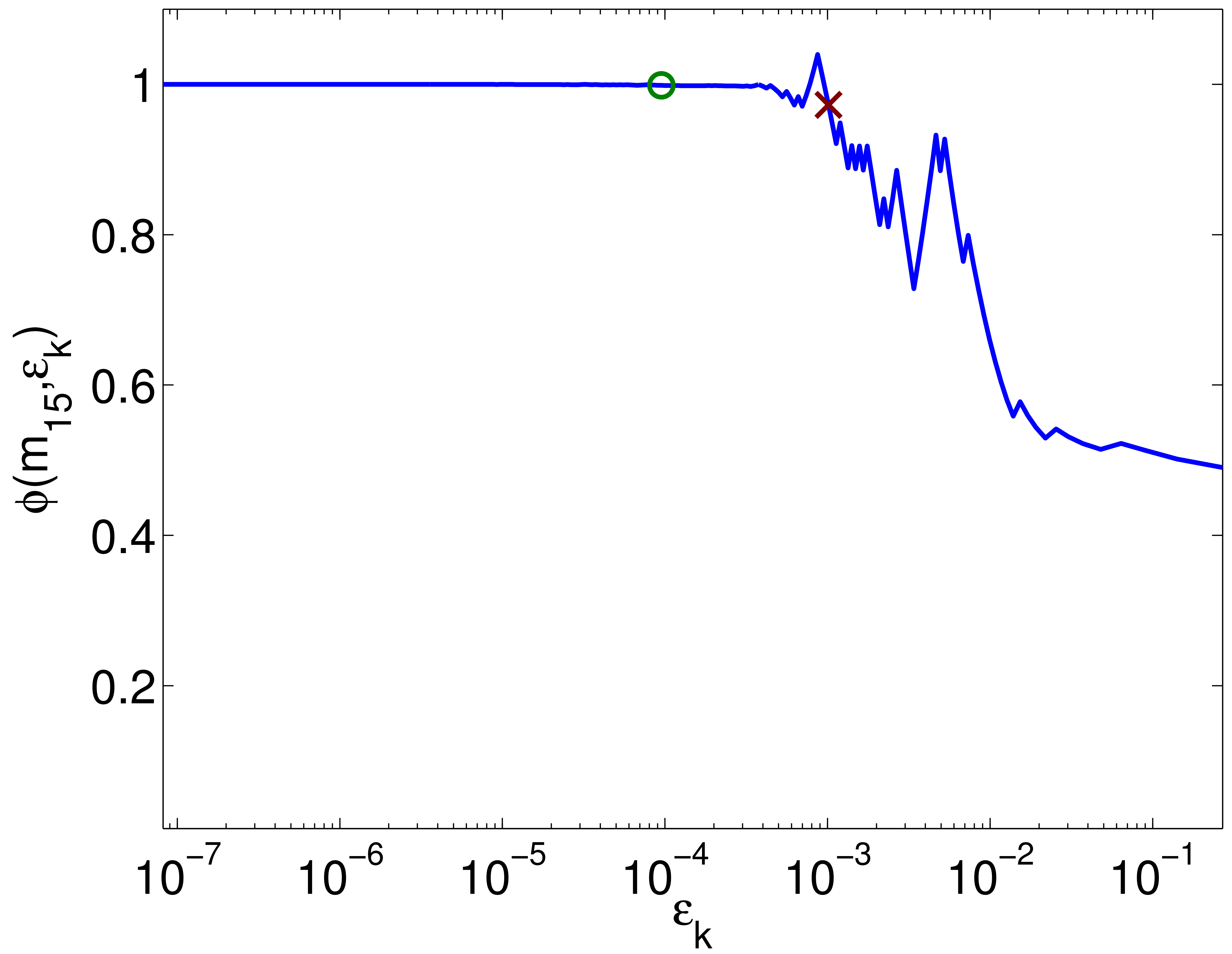
o - relative residual at 10^{-4}
x - relative residual at 10^{-3}



Misfit as a function of ε_k for the **10th** iteration of lBFGS

o - relative residual at 10^{-4}
x - relative residual at 10^{-3}

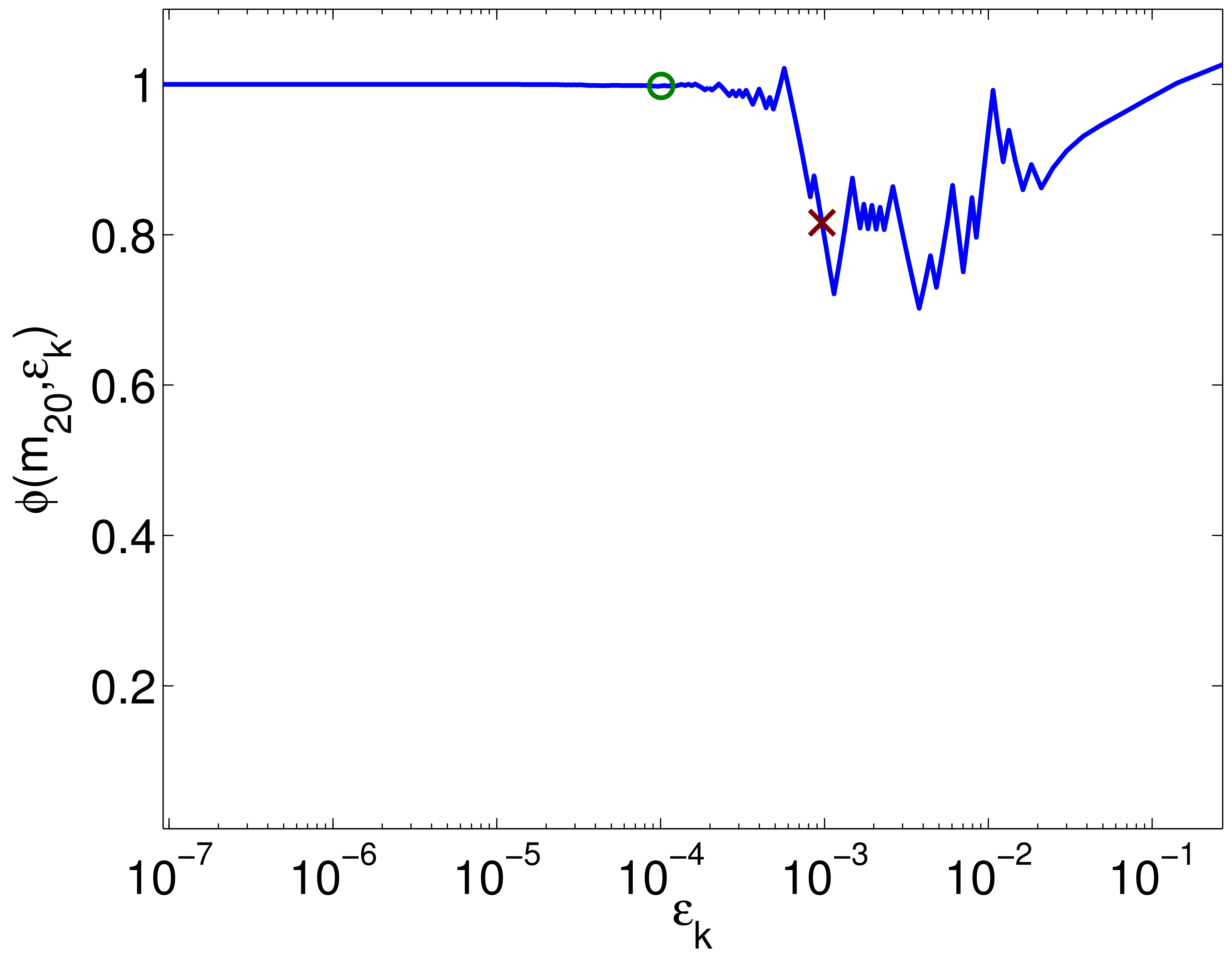
Misfit as a function of ε_k for the **15th** iteration of lBFGS



o - relative residual at 10^{-4}

x - relative residual at 10^{-3}

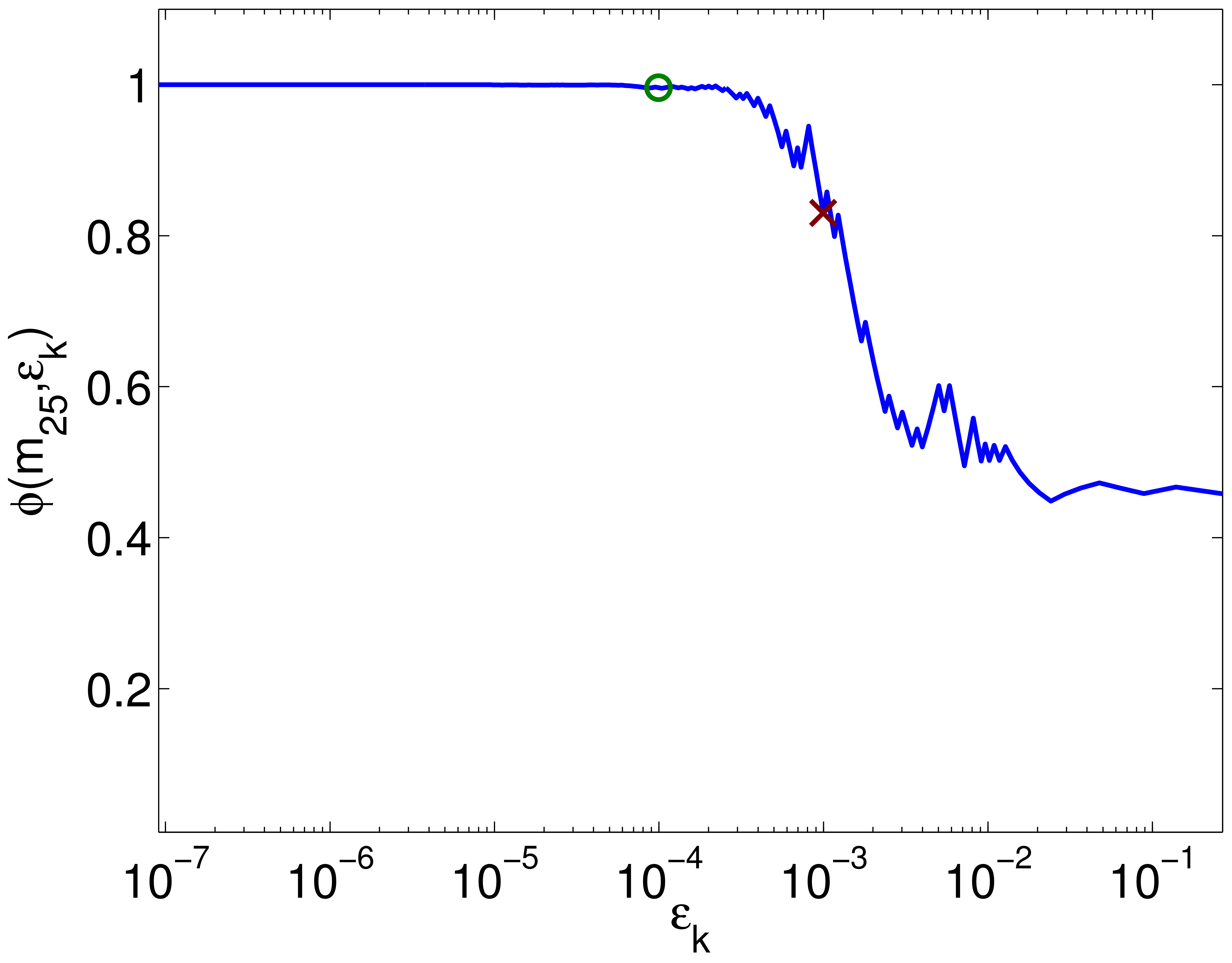
Misfit as a function of ε_k for the **20th** iteration of lBFGS



A Glance Into The Misfit

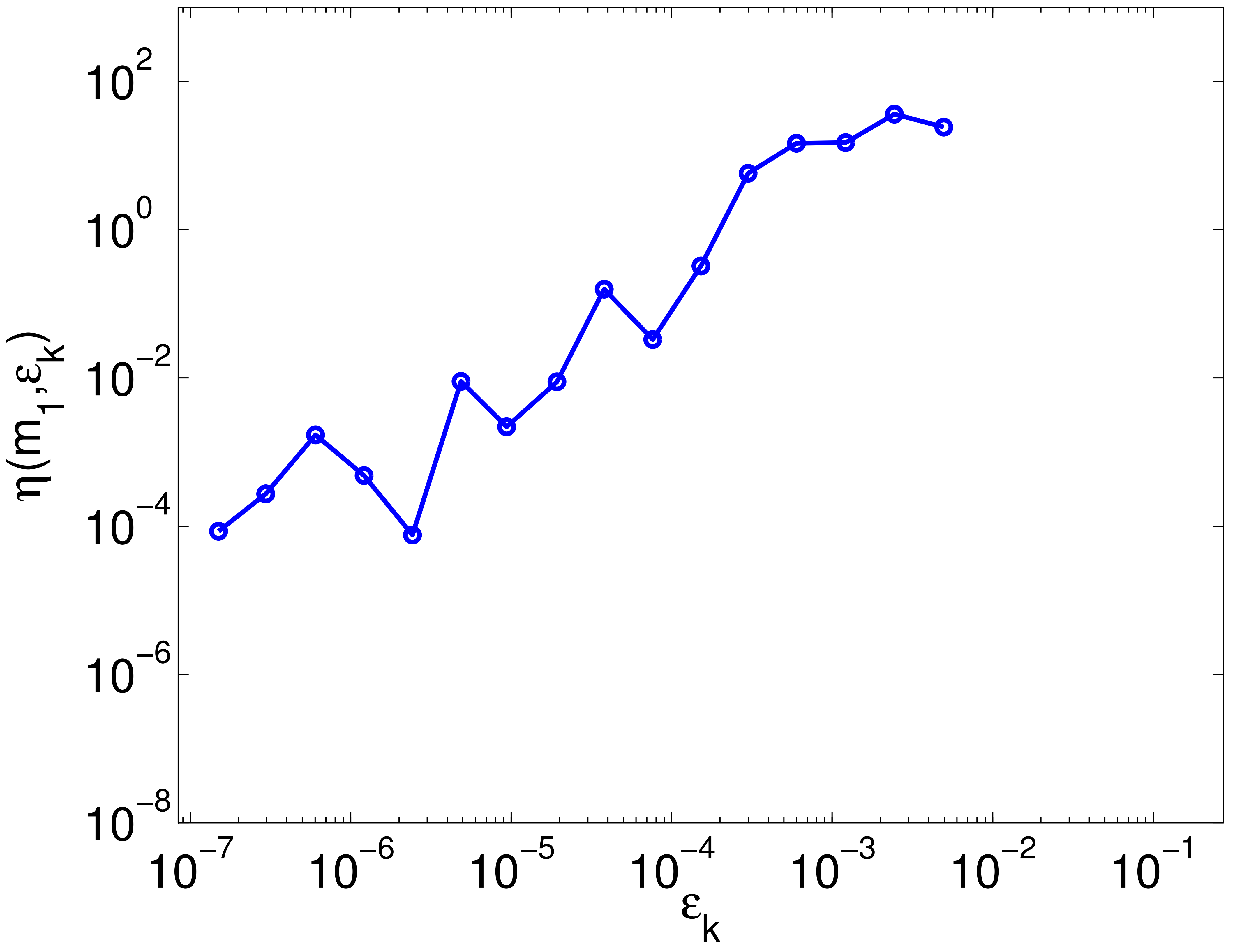
o - relative residual at 10^{-4}
x - relative residual at 10^{-3}

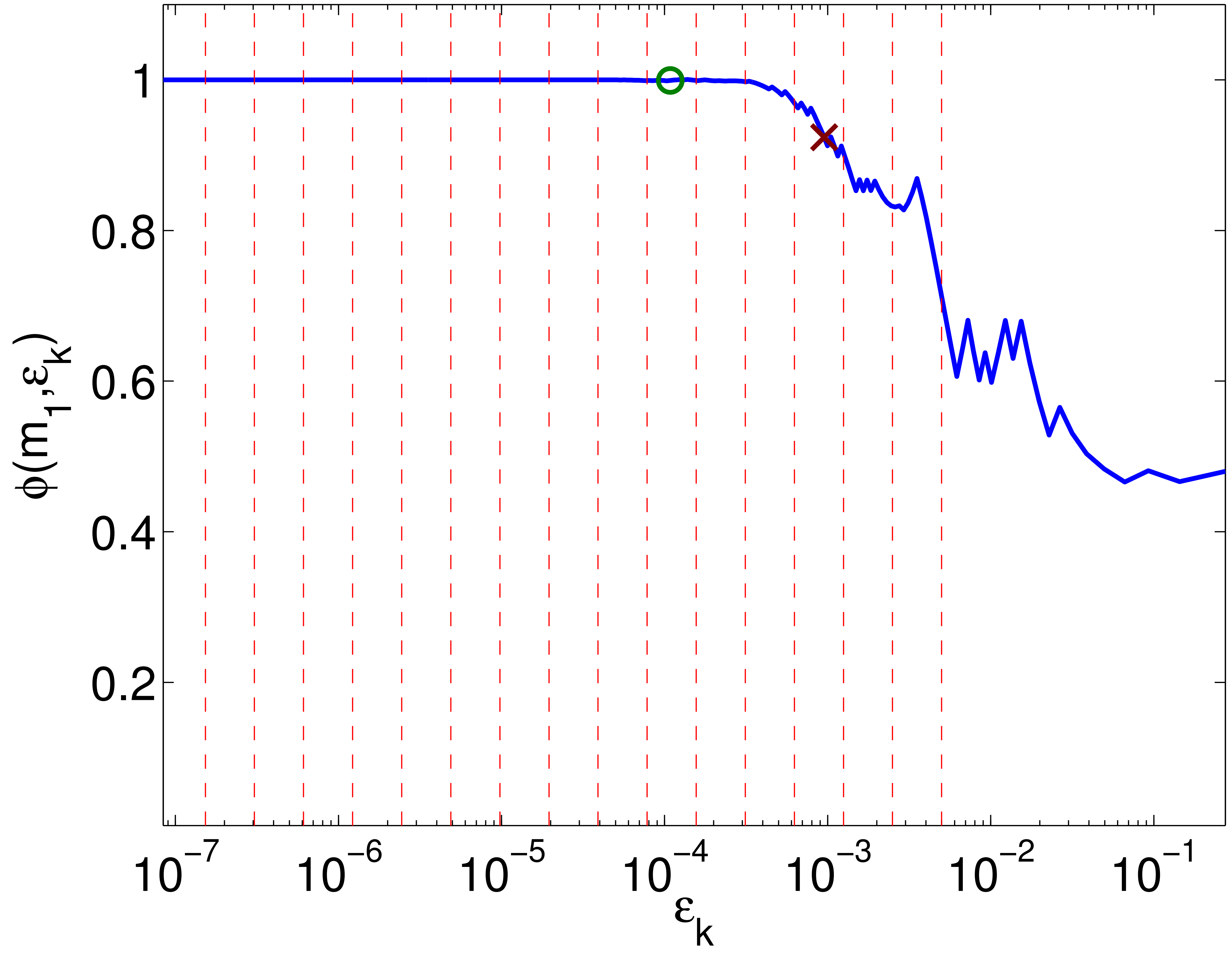
Misfit as a function of ε_k for the **25th** iteration of lBFGS



The α_k Criterion

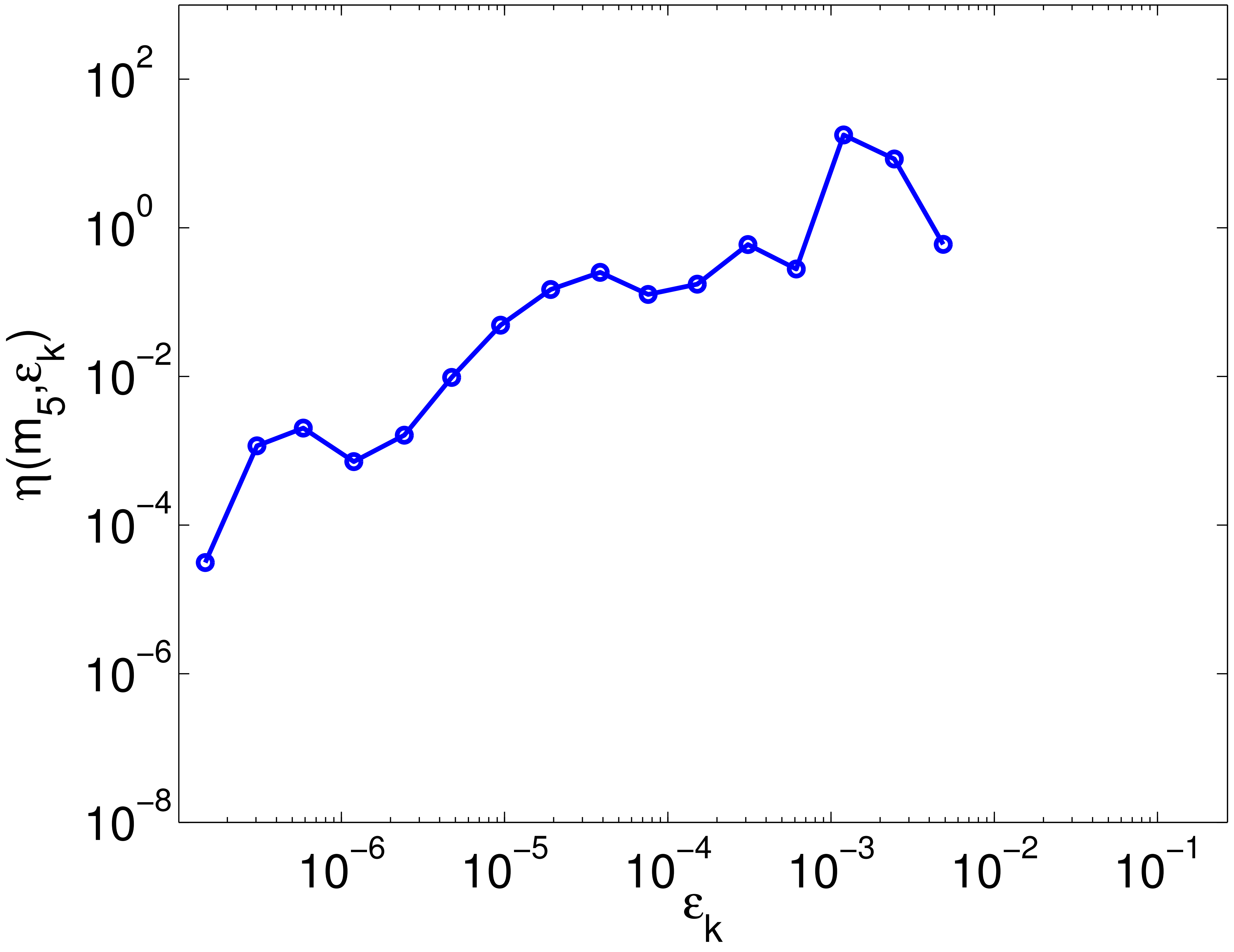
$$\frac{|\Phi(m,\alpha^k\epsilon) - \Phi(m,\alpha^{k+1}\epsilon)|}{\Phi(m,\alpha^{k+1}\epsilon)} \leq \eta \qquad u^i \approx P_r A^{-1}(m) q^i \qquad \frac{\left\|A(m)u^i - q^i\right\|_2}{\left\|q^i\right\|_2} < \alpha^{k+1}\epsilon$$

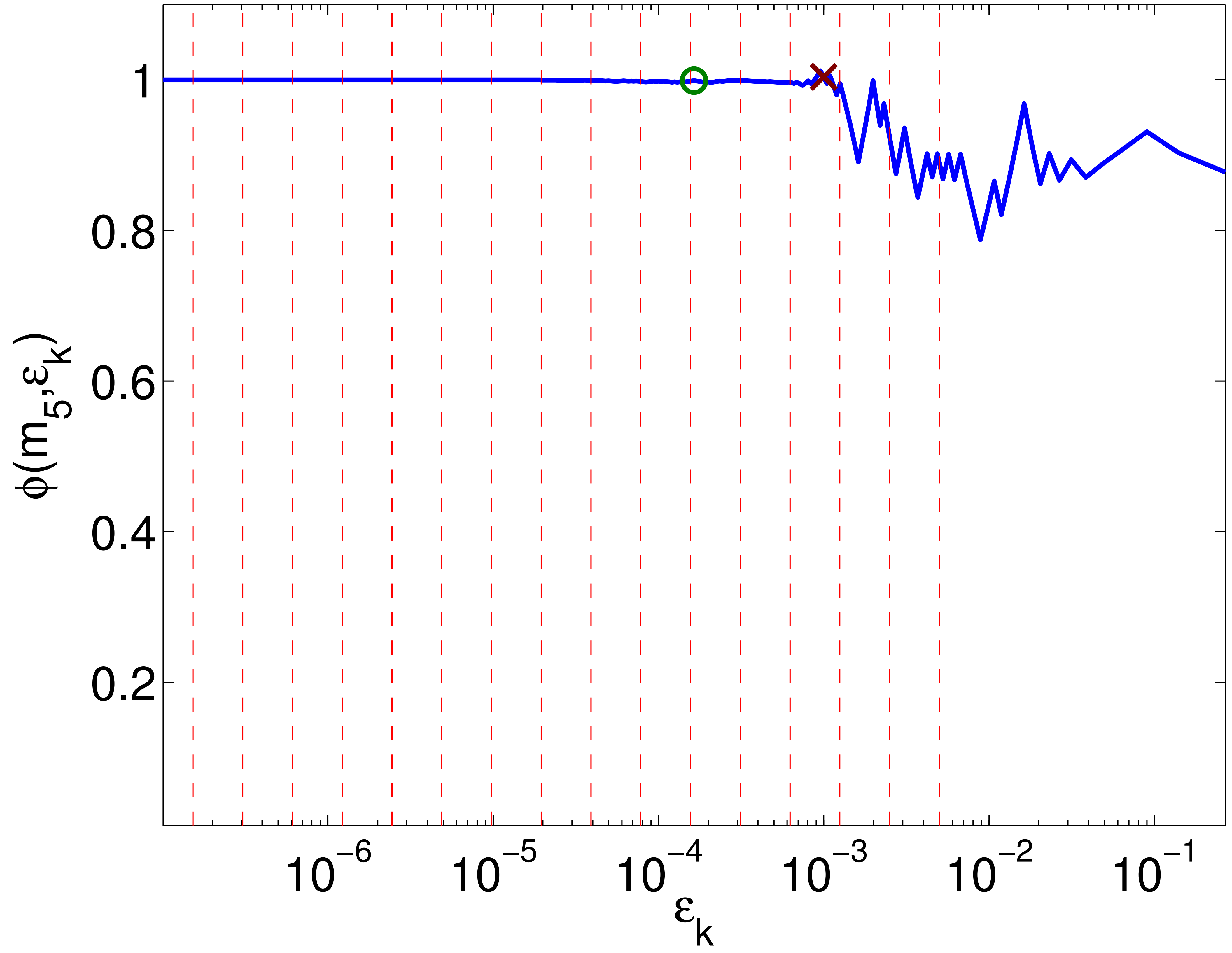




The α_k Criterion

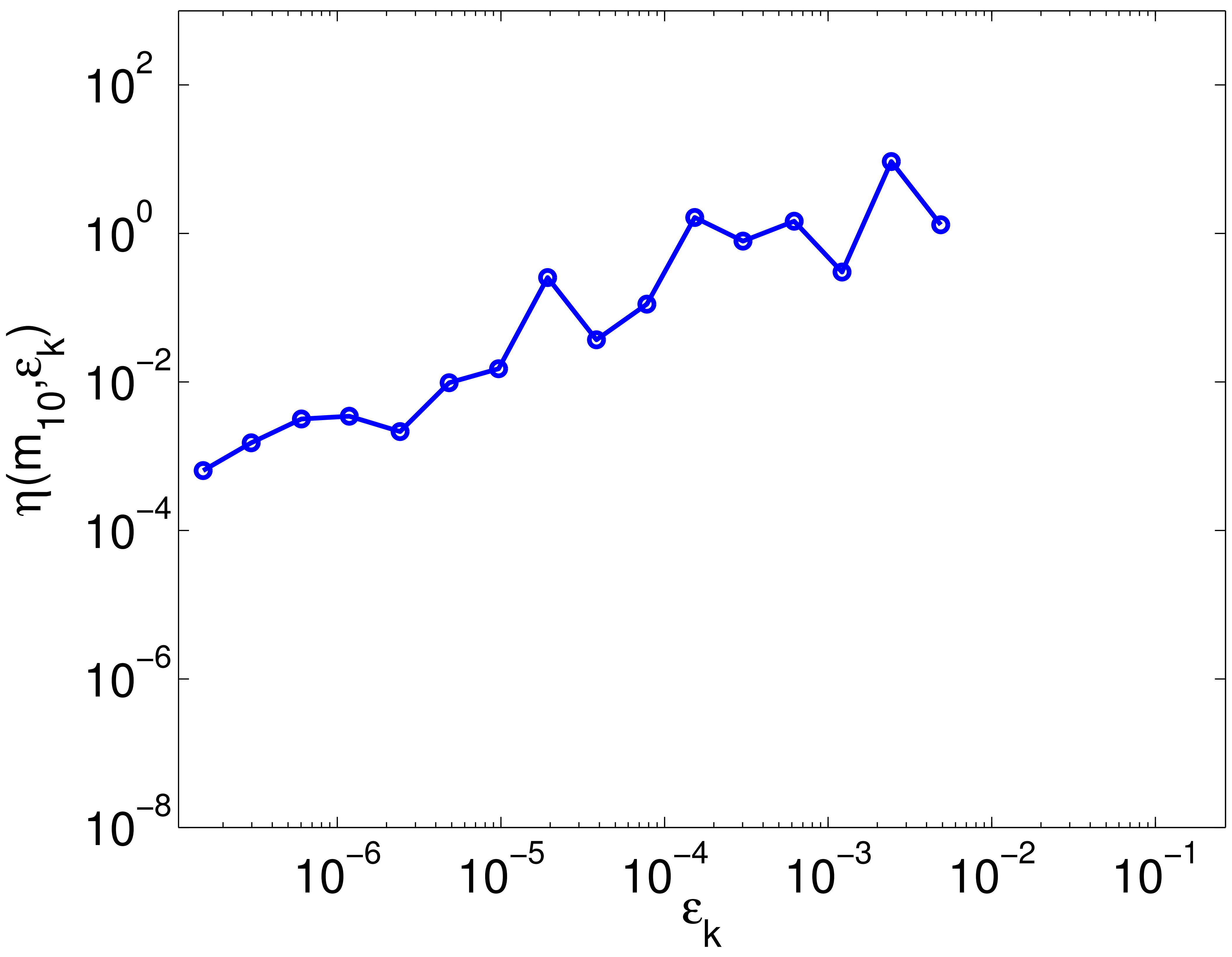
$$\frac{|\Phi(m,\alpha^k\epsilon) - \Phi(m,\alpha^{k+1}\epsilon)|}{\Phi(m,\alpha^{k+1}\epsilon)} \leq \eta \quad u^i \approx P_r A^{-1}(m) q^i \quad \frac{\left\|A(m)u^i - q^i\right\|_2}{\left\|q^i\right\|_2} < \alpha^{k+1}\epsilon$$

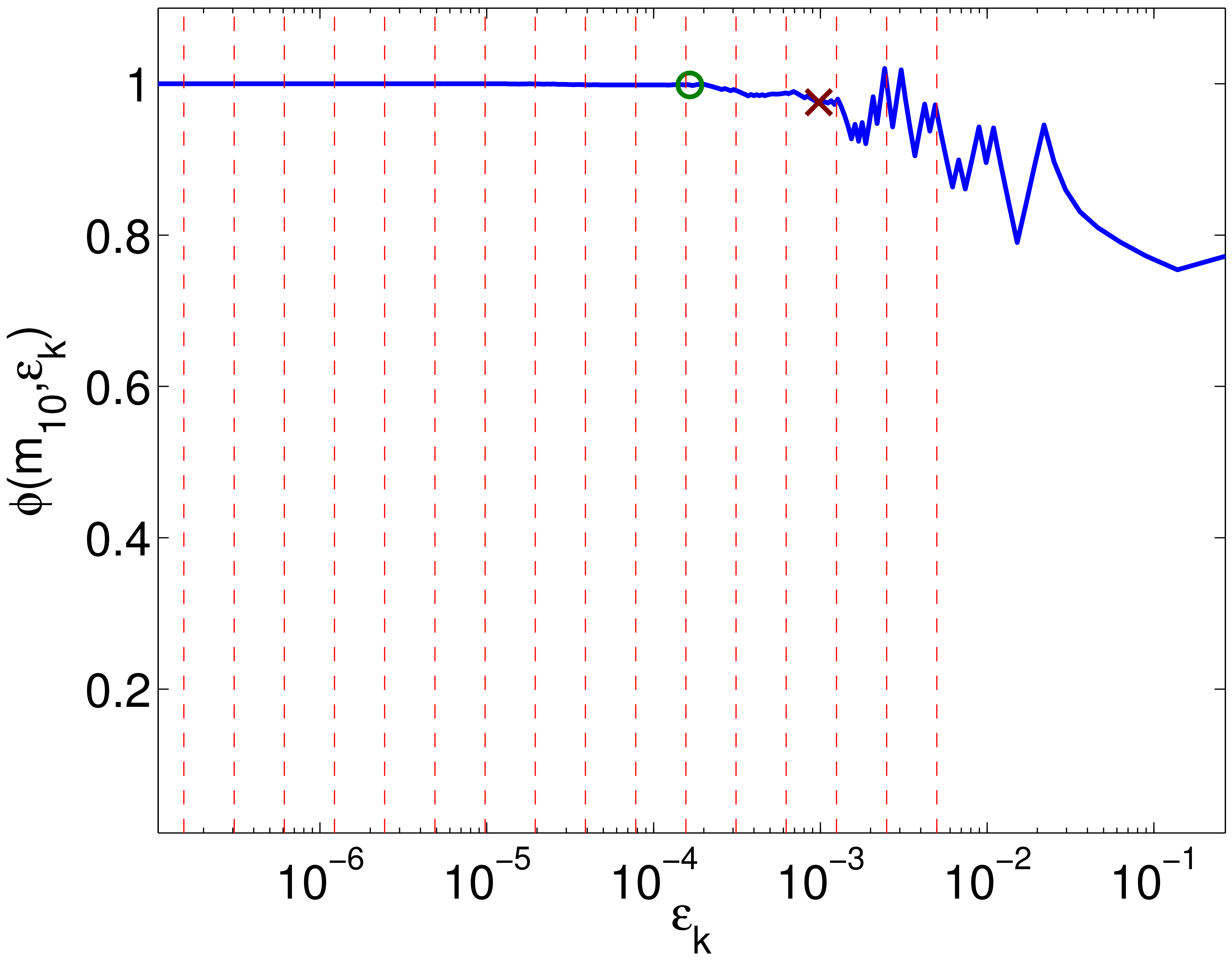




The α_k Criterion

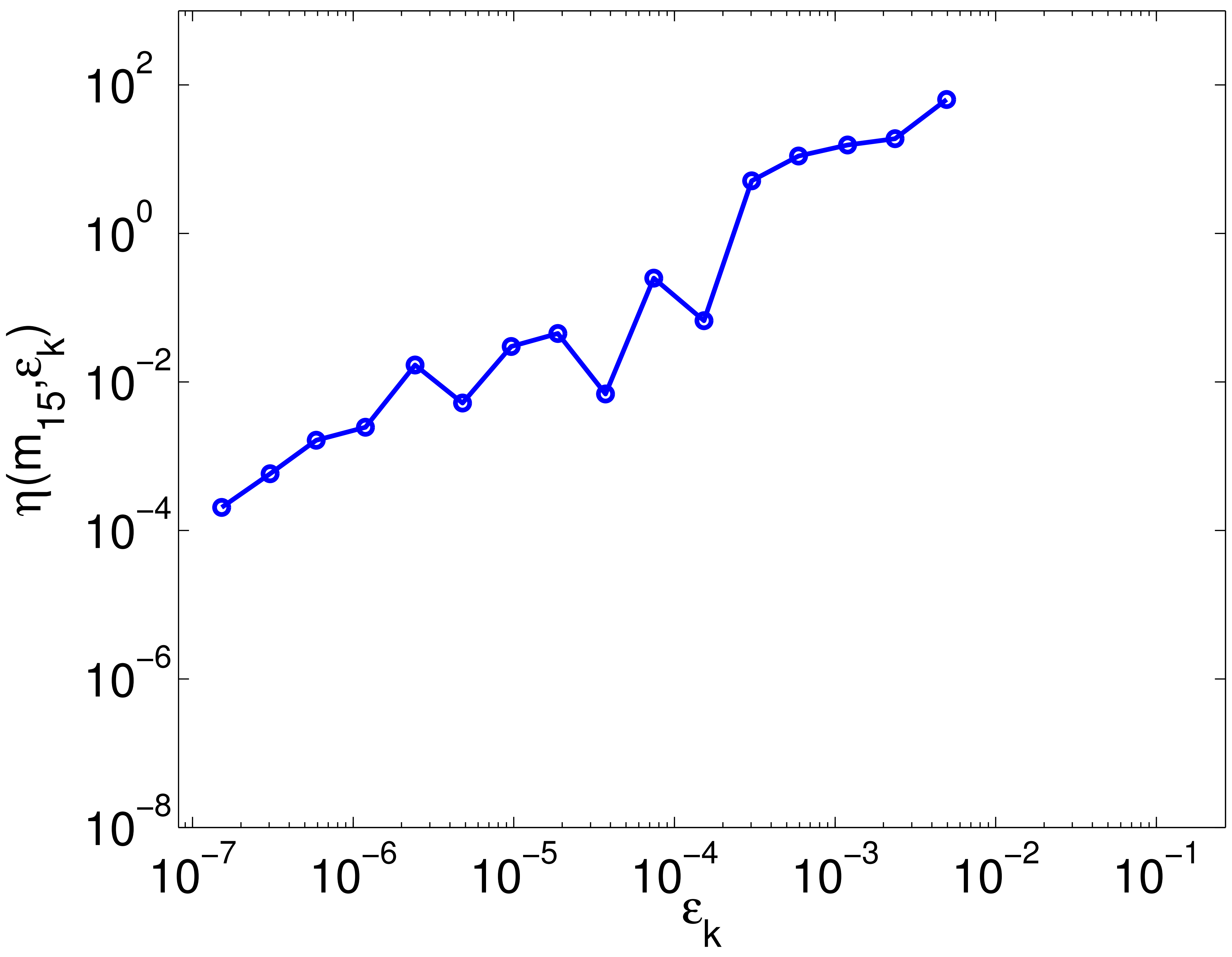
$$\frac{|\Phi(m,\alpha^k\epsilon)-\Phi(m,\alpha^{k+1}\epsilon)|}{\Phi(m,\alpha^{k+1}\epsilon)} \leq \eta \quad u^i \approx P_r A^{-1}(m) q^i \quad \frac{\left\|A(m)u^i - q^i\right\|_2}{\left\|q^i\right\|_2} < \alpha^{k+1}\epsilon$$

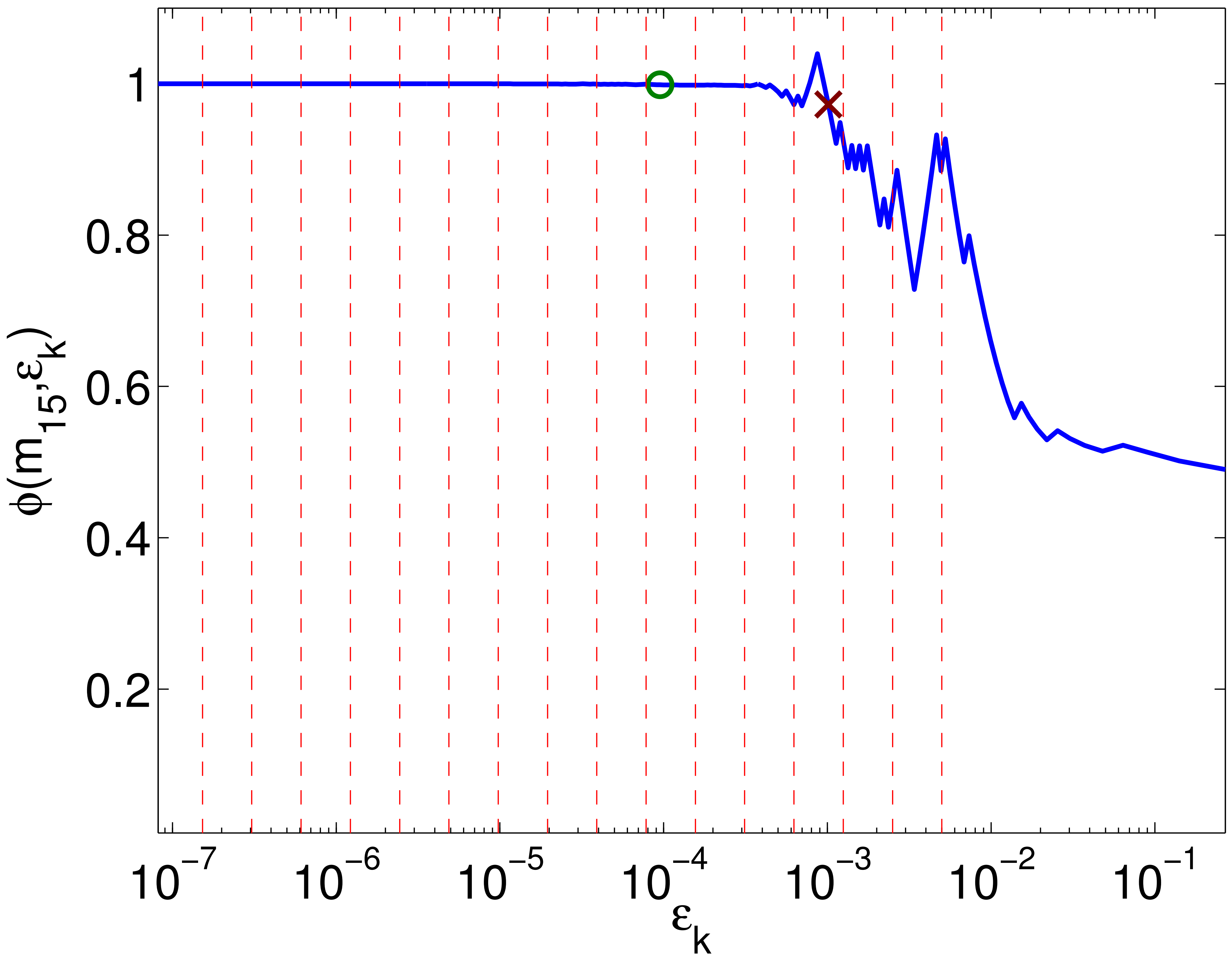




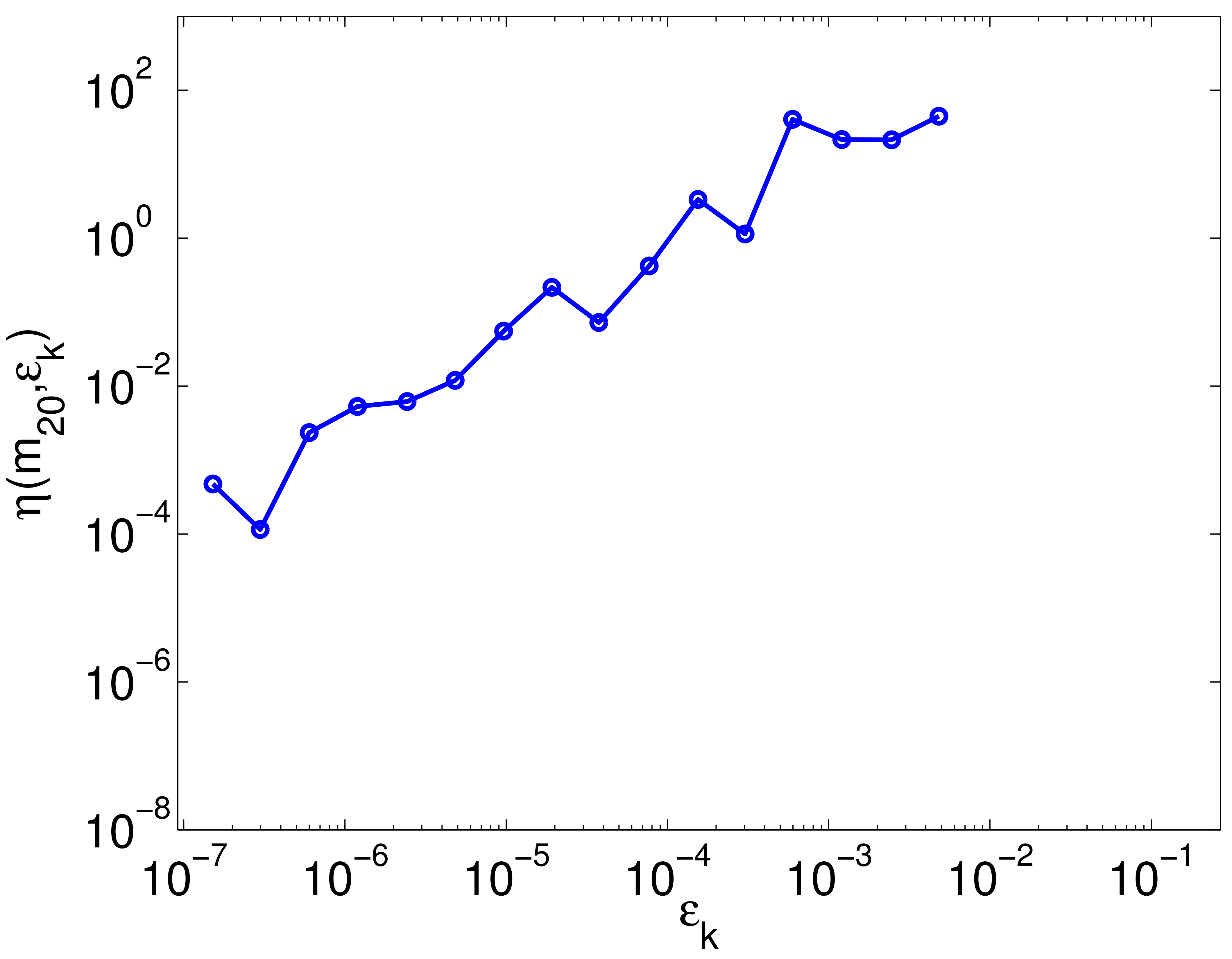
The α_k Criterion

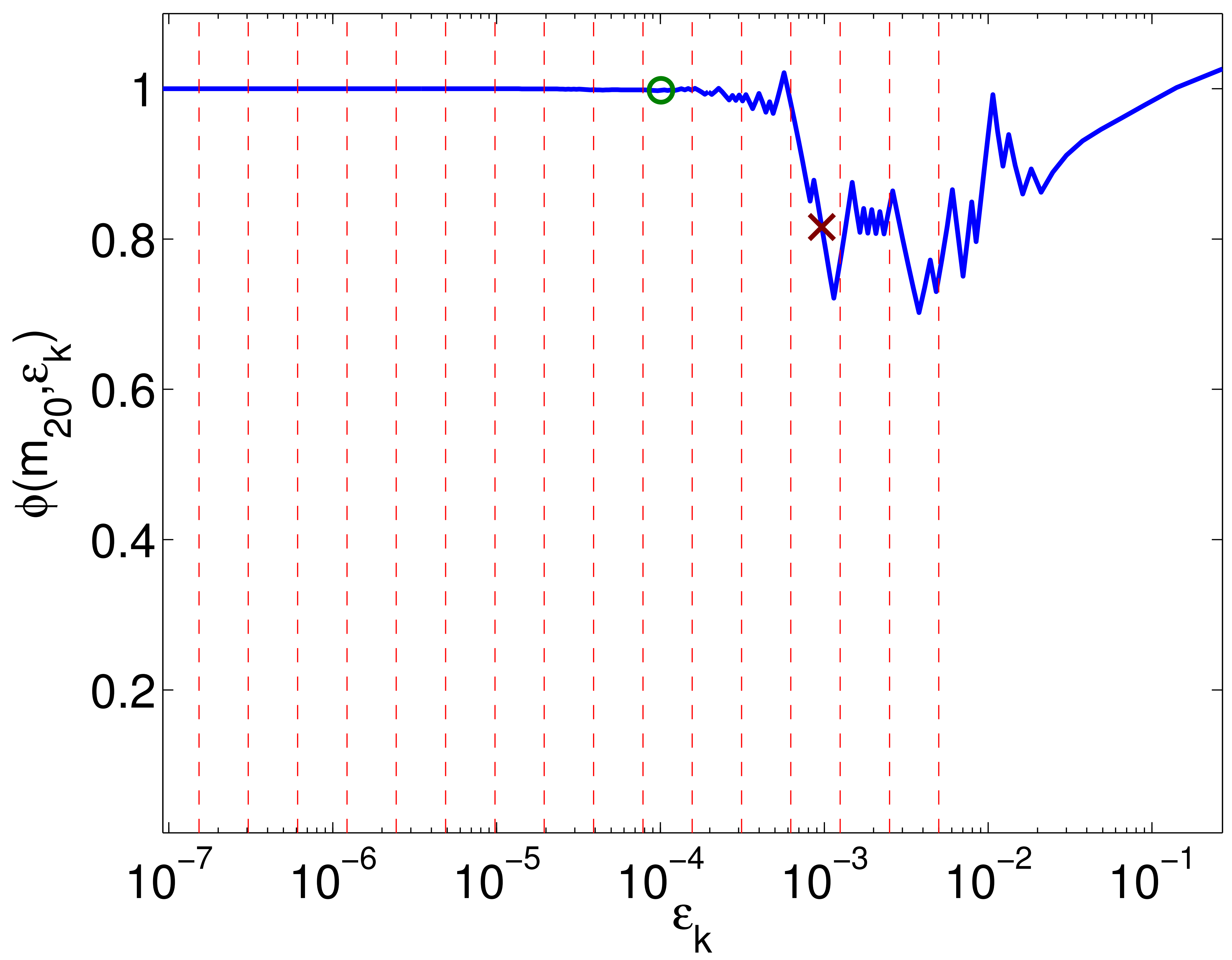
$$\frac{|\Phi(m,\alpha^k\epsilon)-\Phi(m,\alpha^{k+1}\epsilon)|}{\Phi(m,\alpha^{k+1}\epsilon)} \leq \eta \qquad u^i \approx P_r A^{-1}(m) q^i \qquad \frac{\left\|A(m)u^i - q^i\right\|_2}{\left\|q^i\right\|_2} < \alpha^{k+1}\epsilon$$





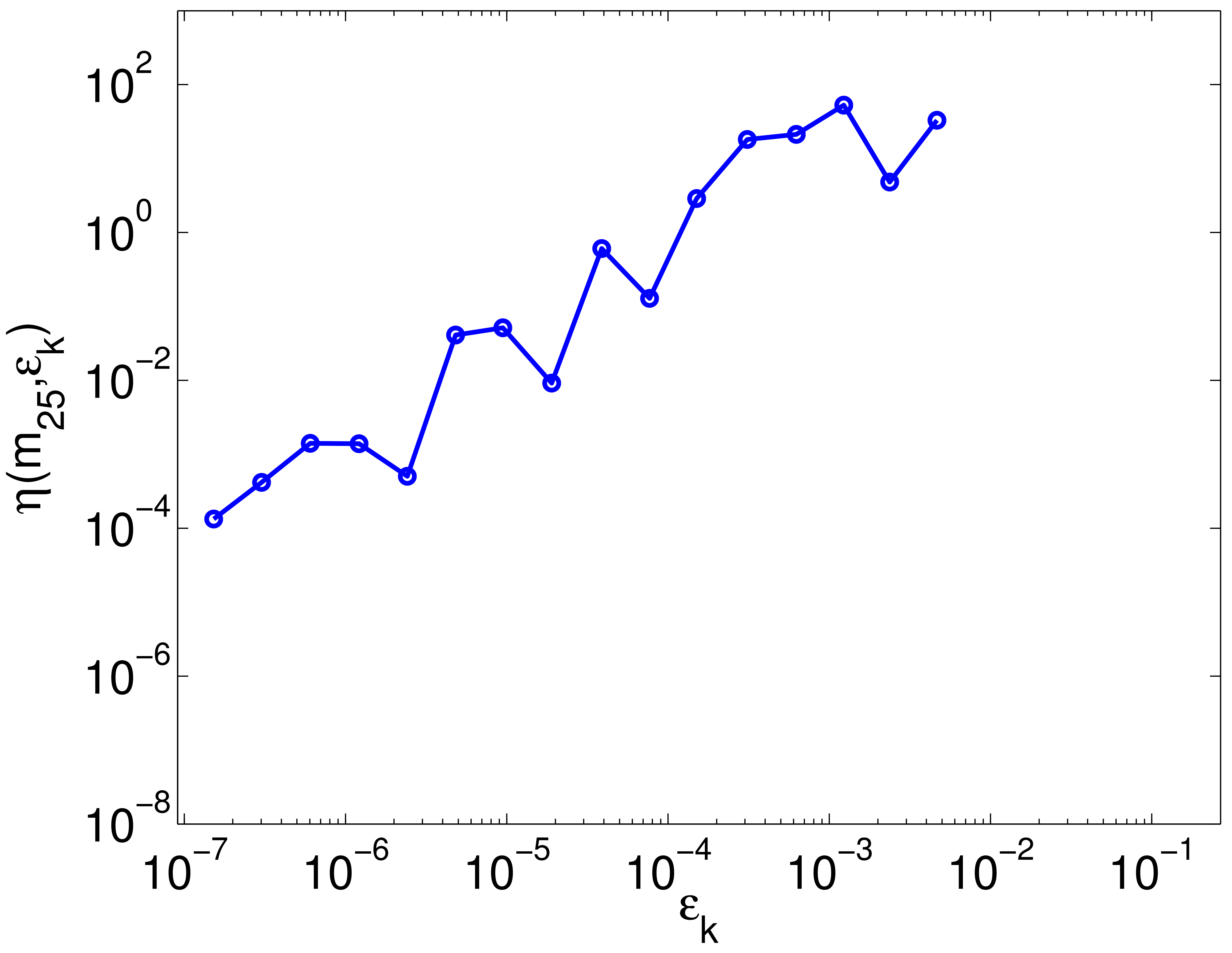
$$\frac{|\Phi(m, \alpha^k \epsilon) - \Phi(m, \alpha^{k+1}\epsilon)|}{\Phi(m, \alpha^{k+1}\epsilon)} \leq \eta \quad u^i \approx P_r A^{-1}(m) q^i \quad \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2} < \alpha^{k+1}\epsilon$$

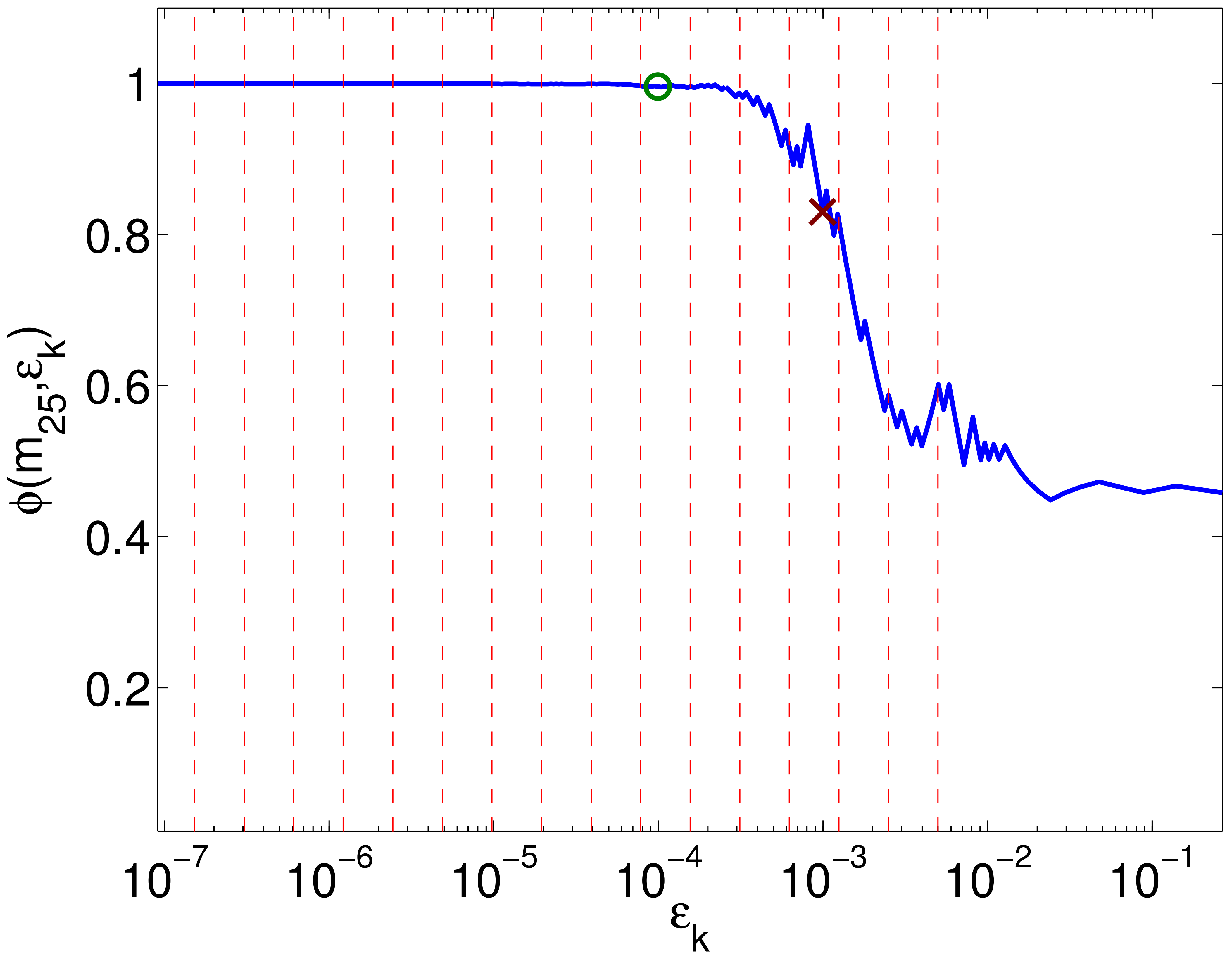




The α_k Criterion

$$\frac{|\Phi(m, \alpha^k \epsilon) - \Phi(m, \alpha^{k+1}\epsilon)|}{\Phi(m, \alpha^{k+1}\epsilon)} \leq \eta \quad u^i \approx P_r A^{-1}(m) q^i \quad \frac{\left\| A(m) u^i - q^i \right\|_2}{\left\| q^i \right\|_2} < \alpha^{k+1}\epsilon$$

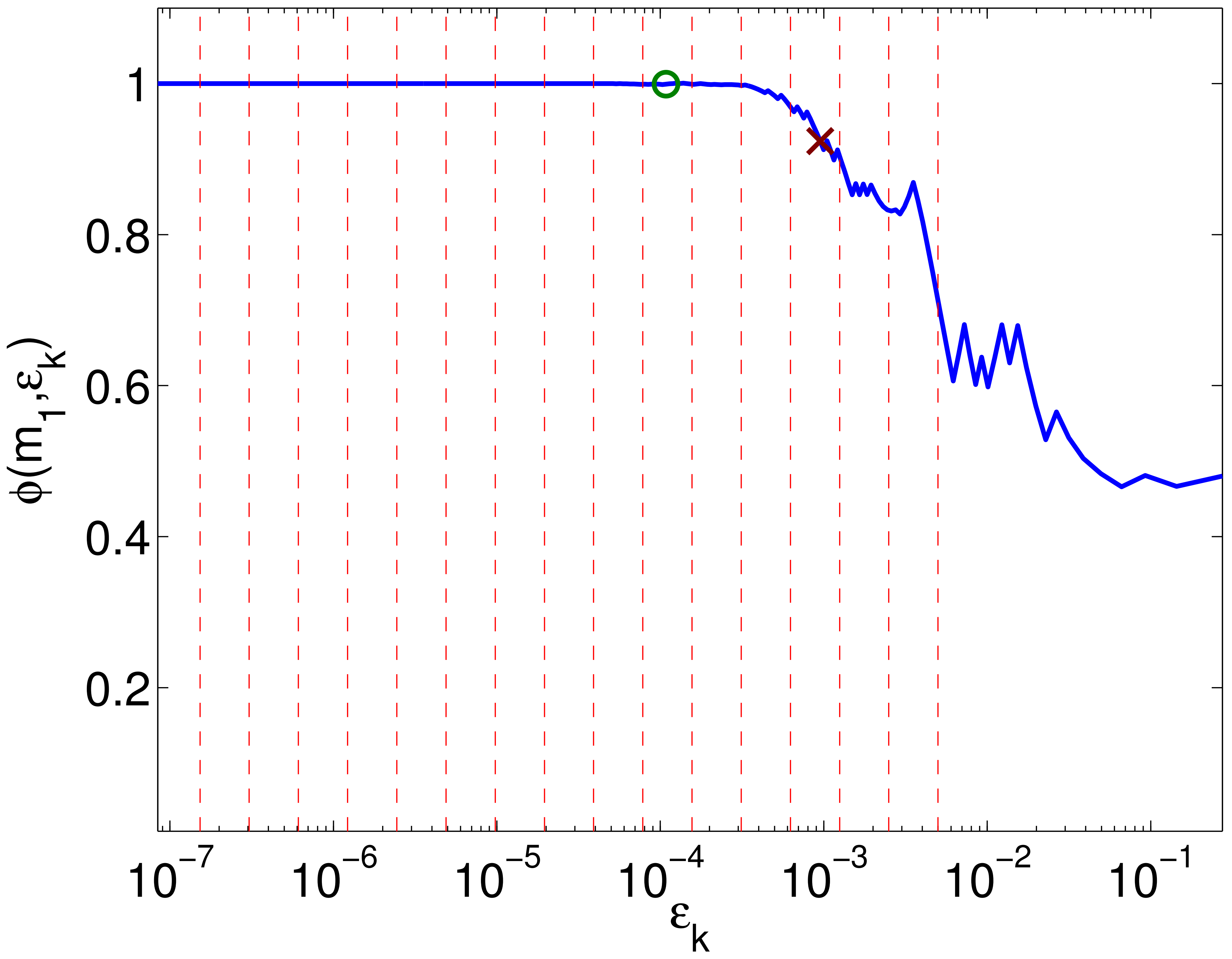




Finer Sampling

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{\Phi(m,\varepsilon_{k+1})} \leq \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

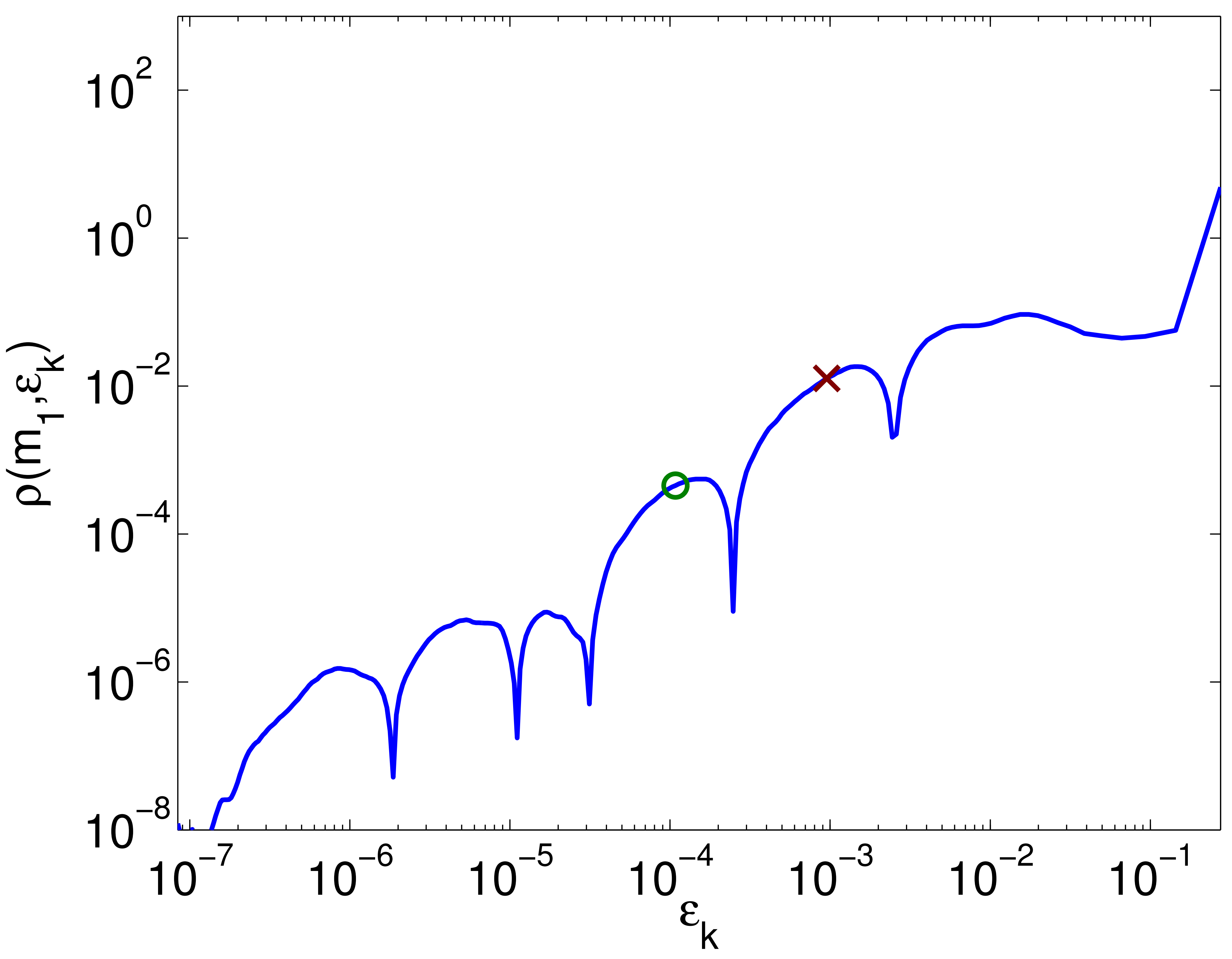
$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

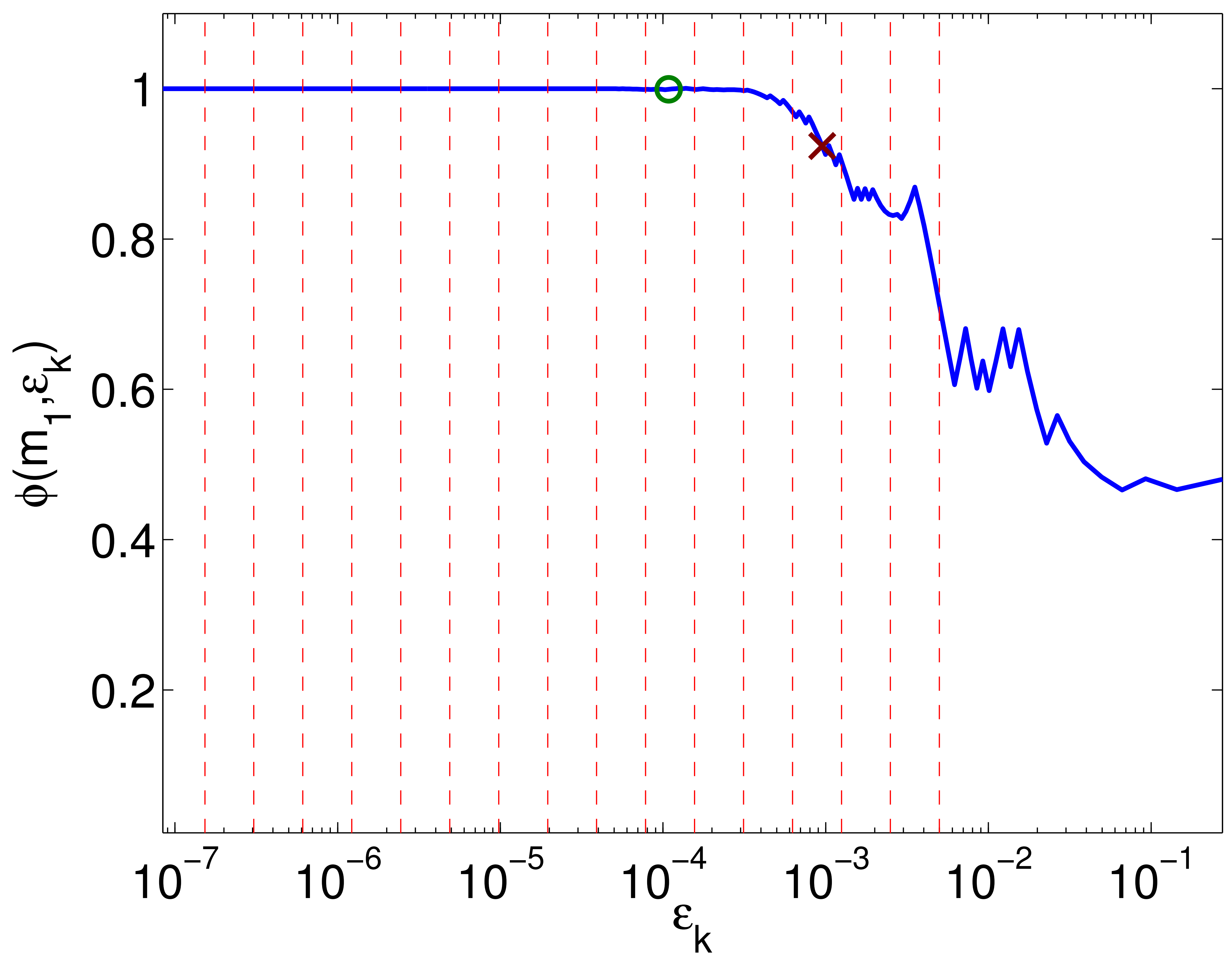


Finer Sampling

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{\Phi(m,\varepsilon_{k+1})} \leq \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

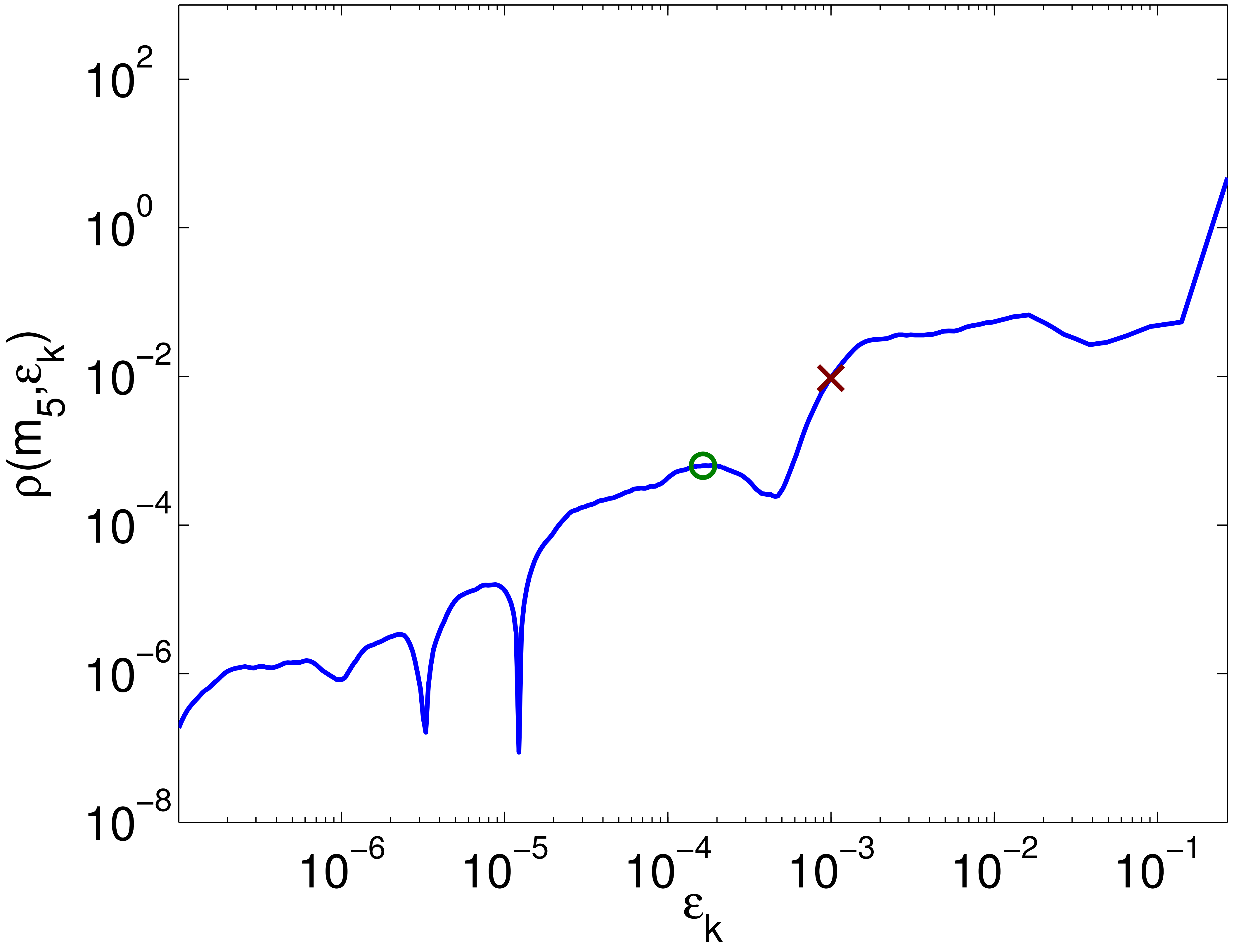


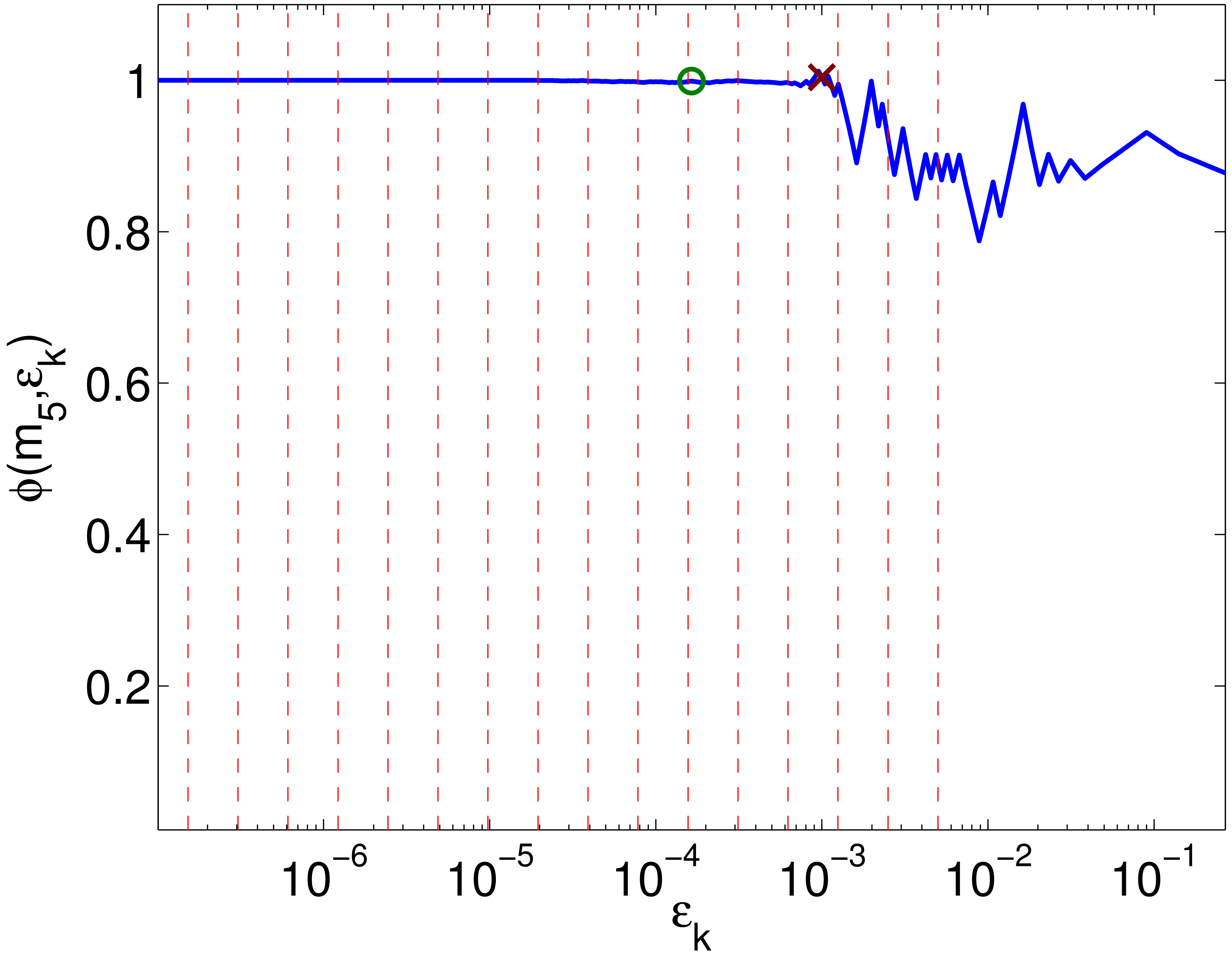


Finer Sampling

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{\Phi(m,\varepsilon_{k+1})} \leq \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

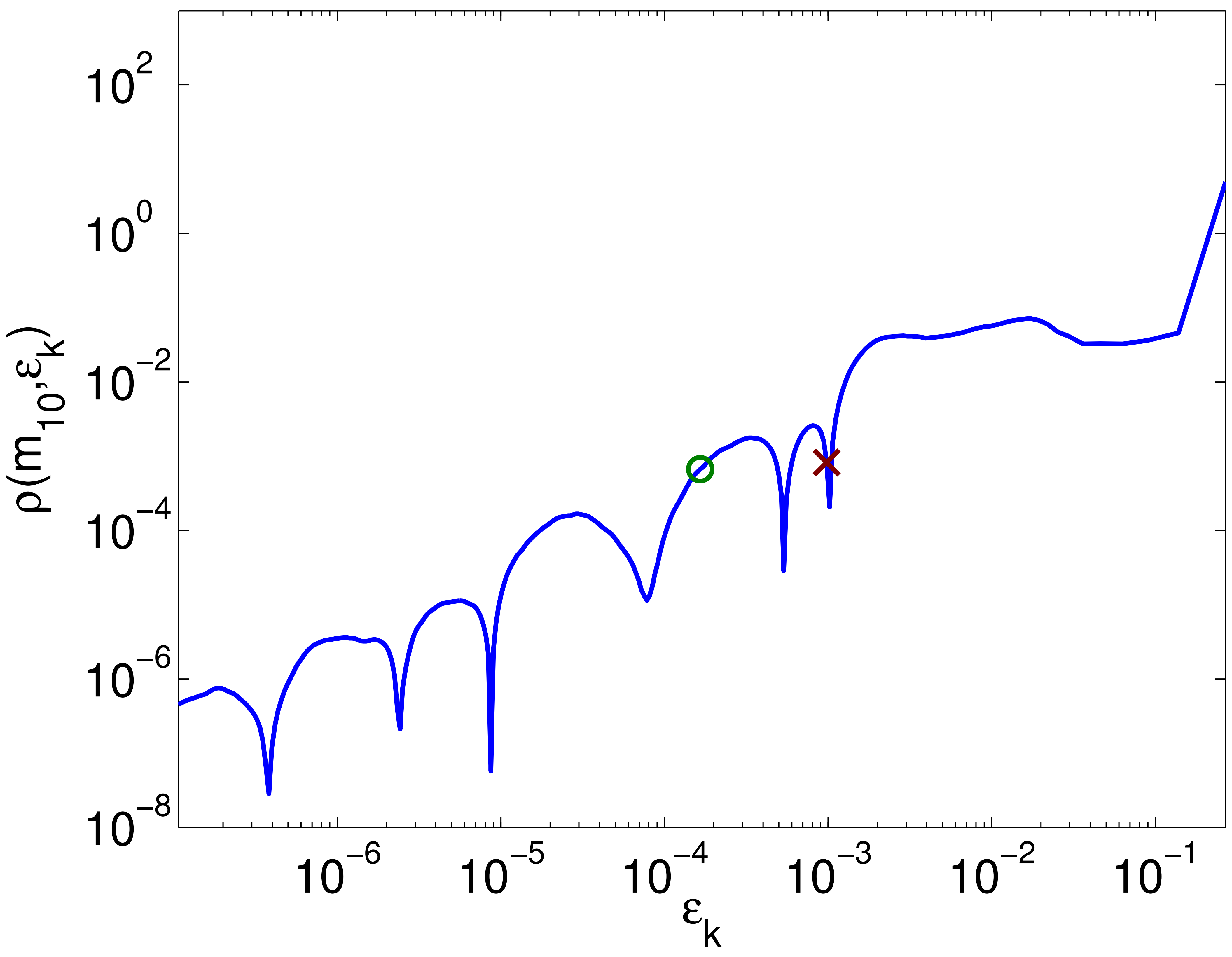


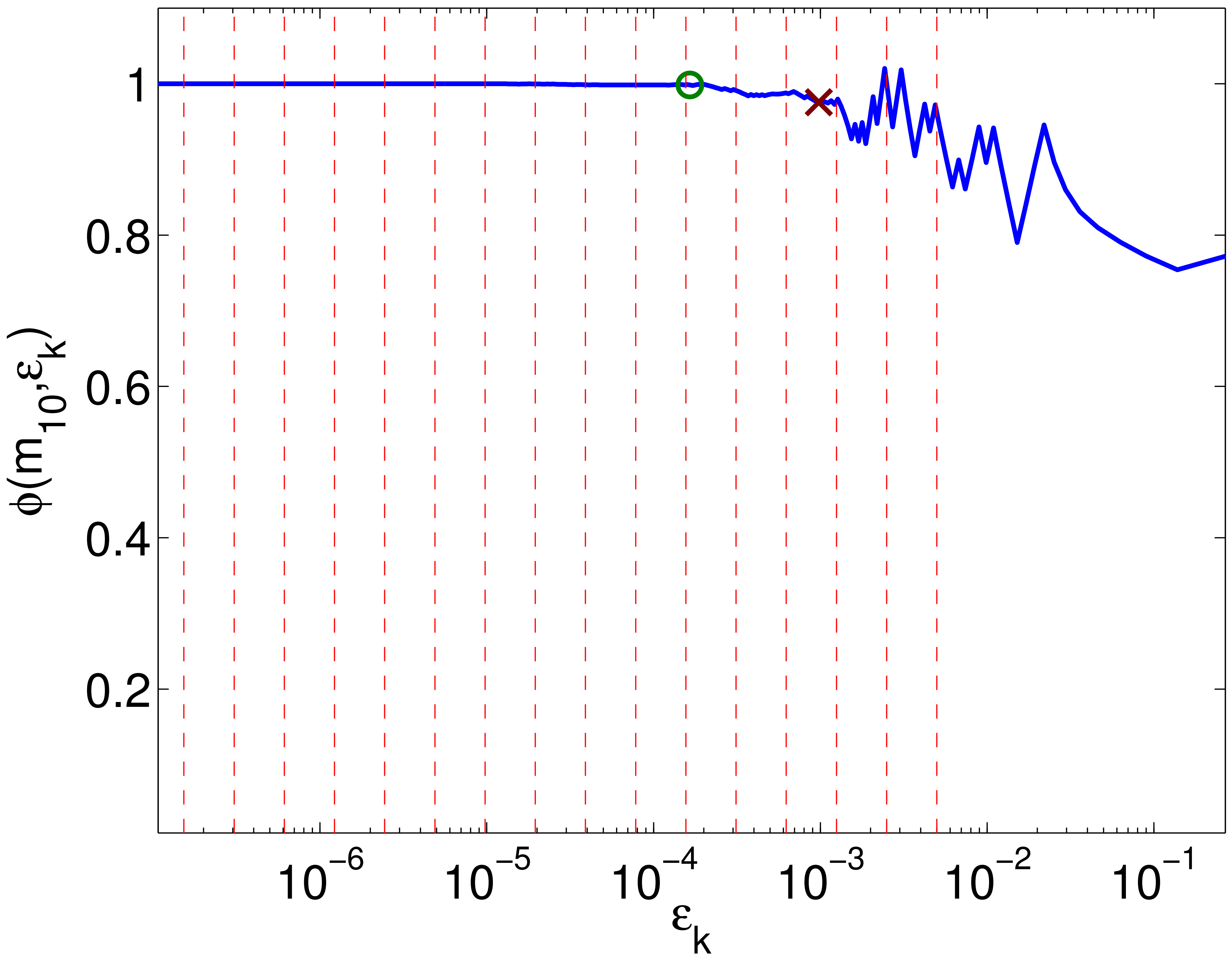


Finer Sampling

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{\Phi(m,\varepsilon_{k+1})} \leq \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

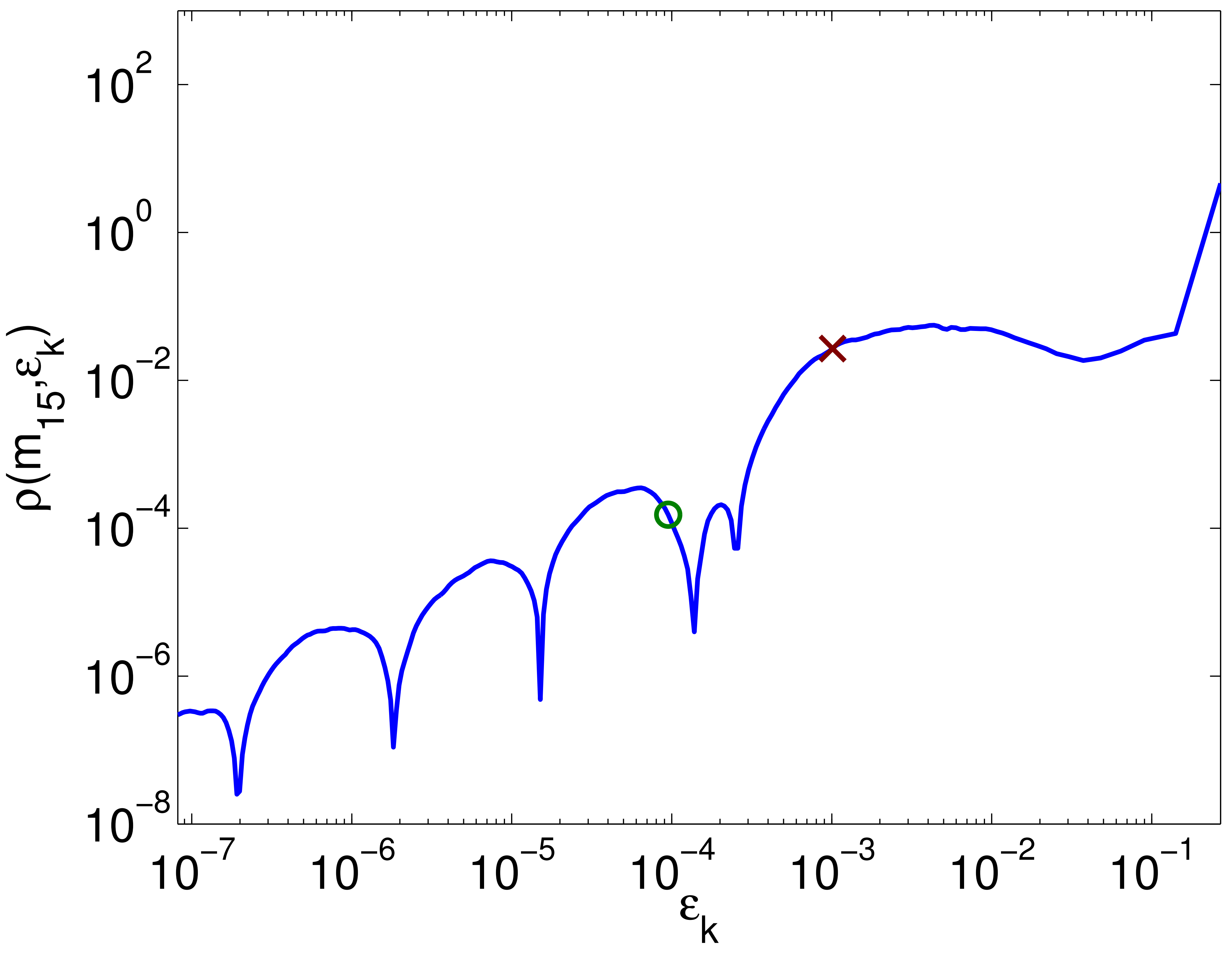


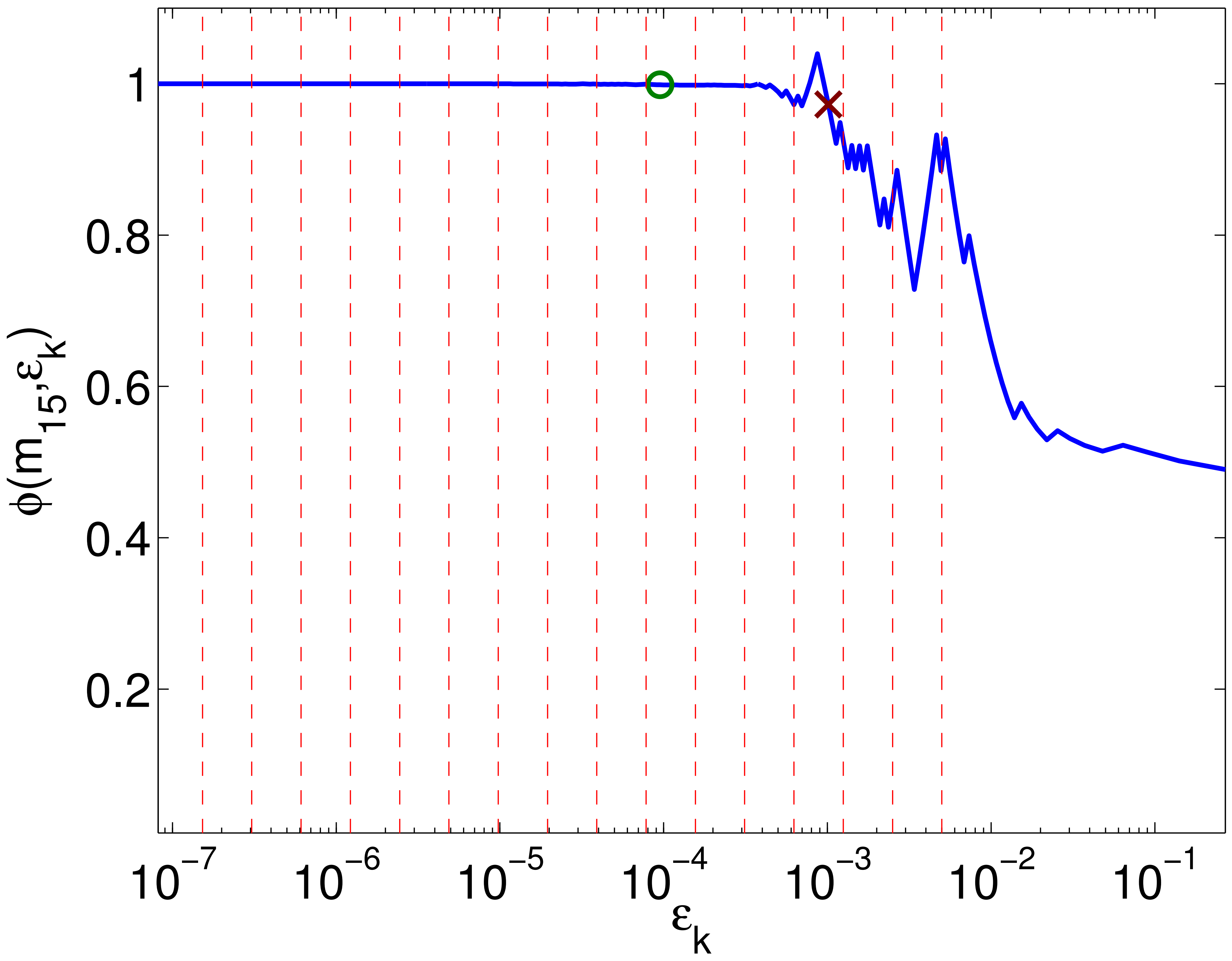


Finer Sampling

$$\frac{|\Phi(m,\varepsilon_k)-\Phi(m,\varepsilon_{k+1})|}{\Phi(m,\varepsilon_{k+1})} \leq \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

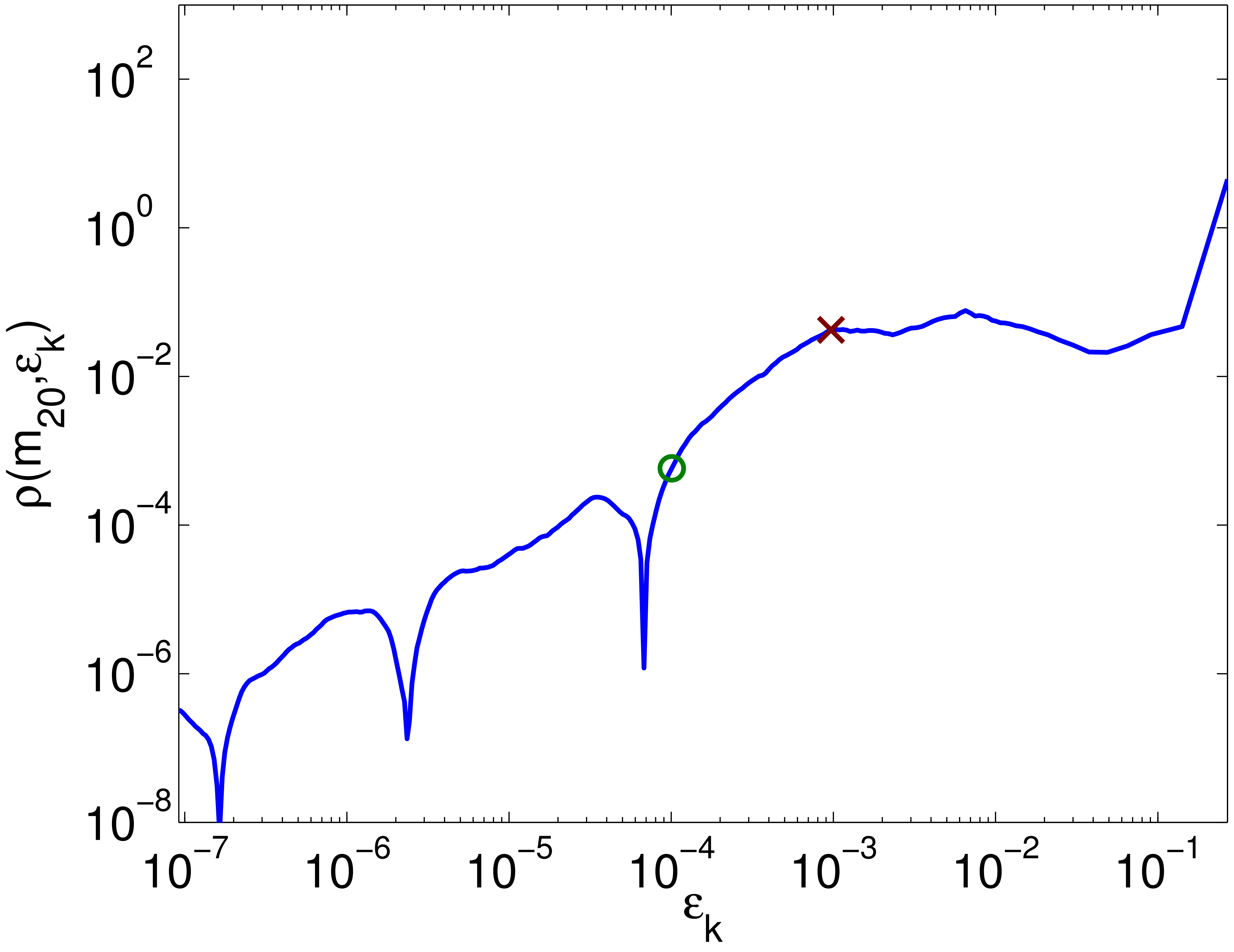


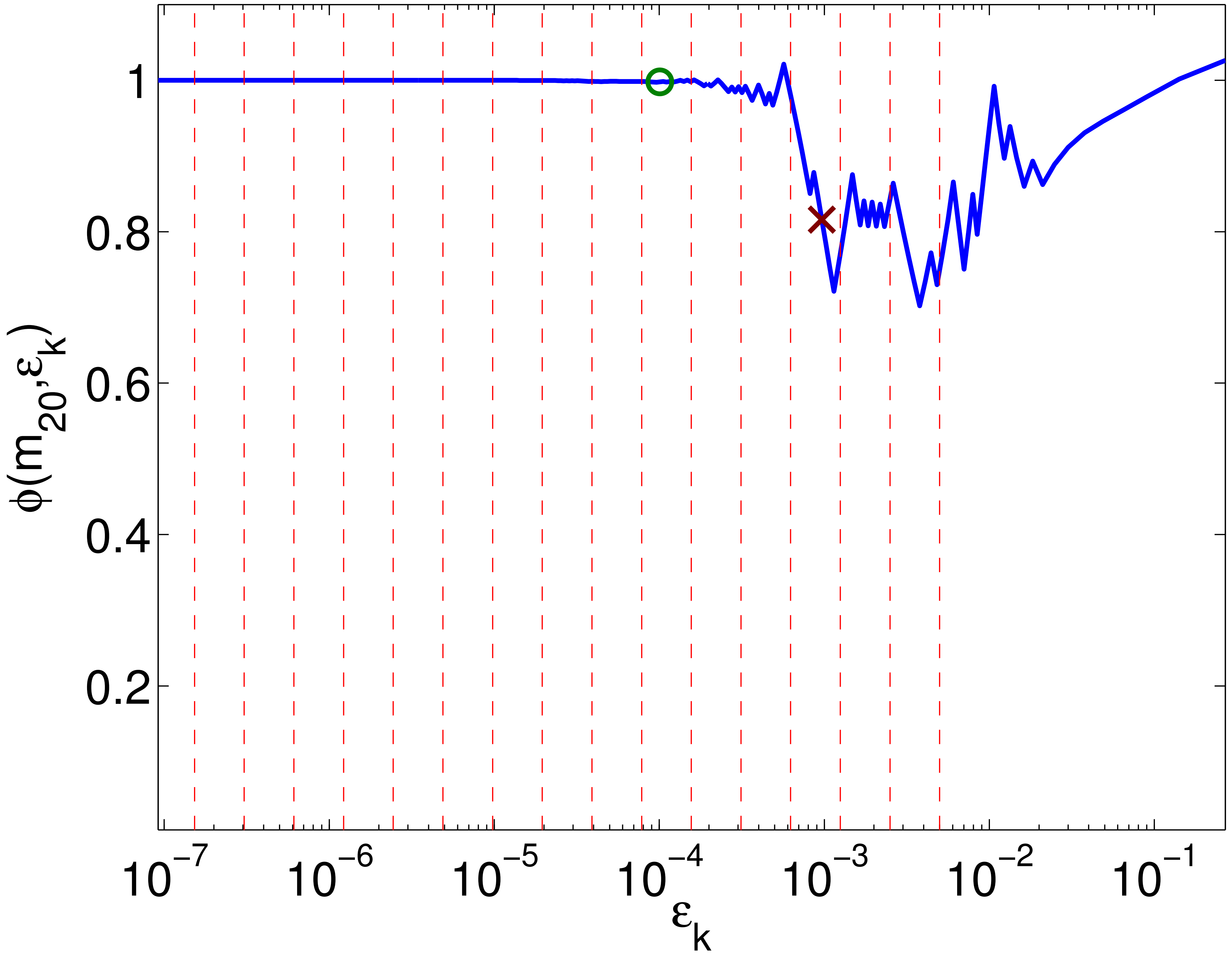


Finer Sampling

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{\Phi(m,\varepsilon_{k+1})} \leq \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

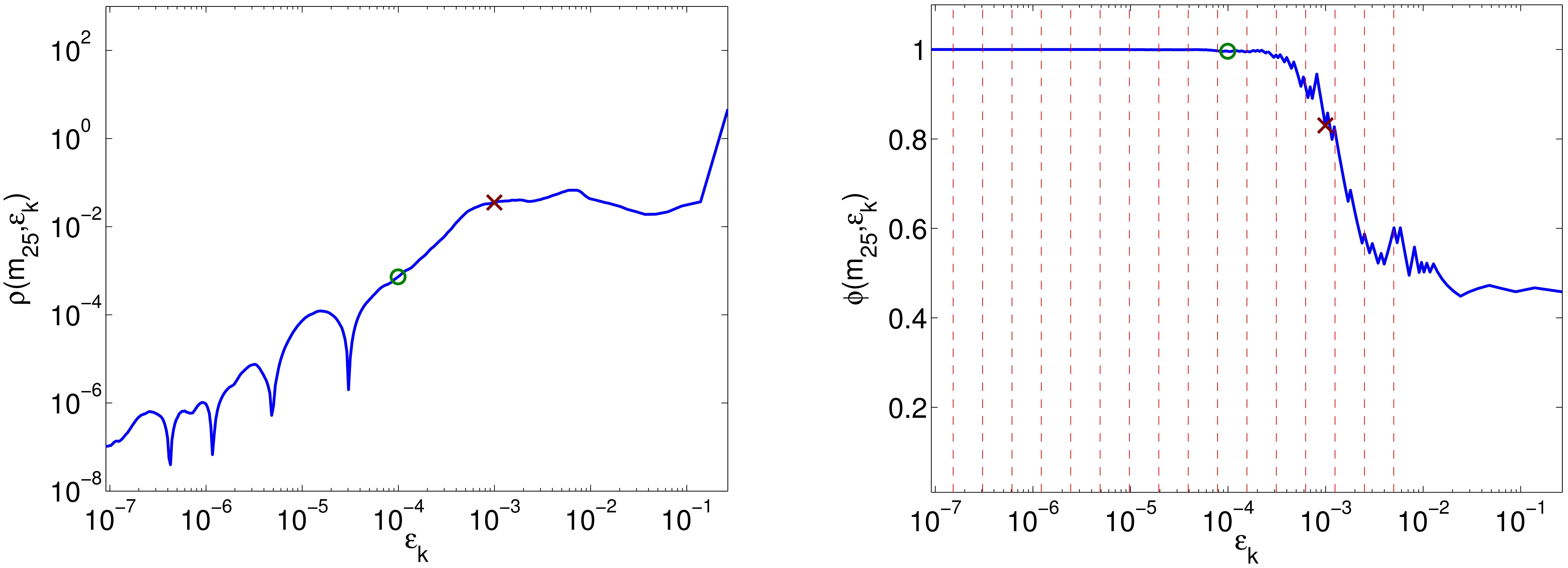
$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$





$$\frac{\left|\Phi(m,\varepsilon_k)-\Phi(m,\varepsilon_{k+1})\right|}{\Phi(m,\varepsilon_{k+1})} \le \eta \frac{\varepsilon_{k+1}}{\varepsilon_k}$$

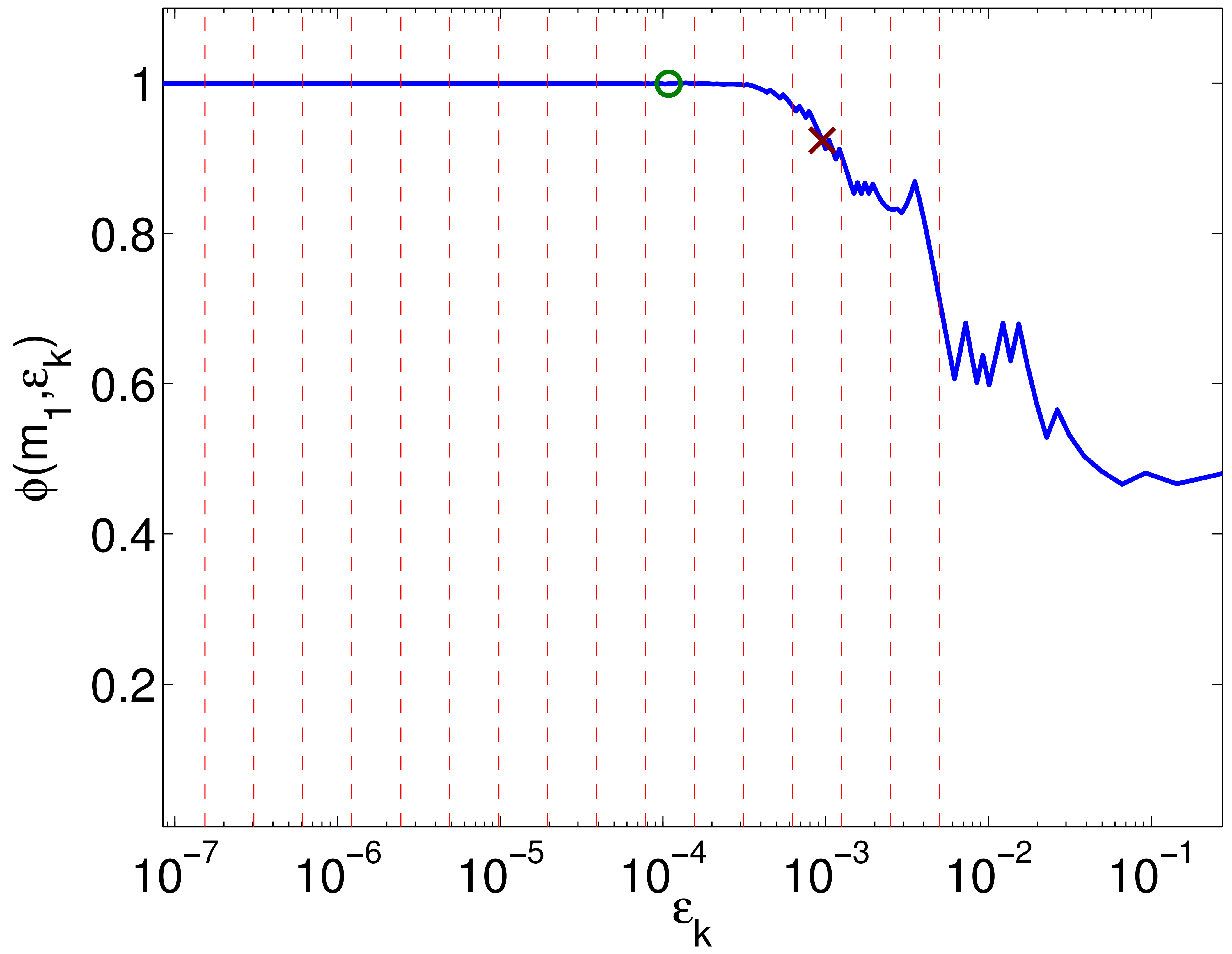
$$\varepsilon_k := \frac{\left\|A(m)u^i - q^i\right\|_2}{\left\|q^i\right\|_2}$$



First Order - The $\delta\varepsilon_k$ Criterion

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

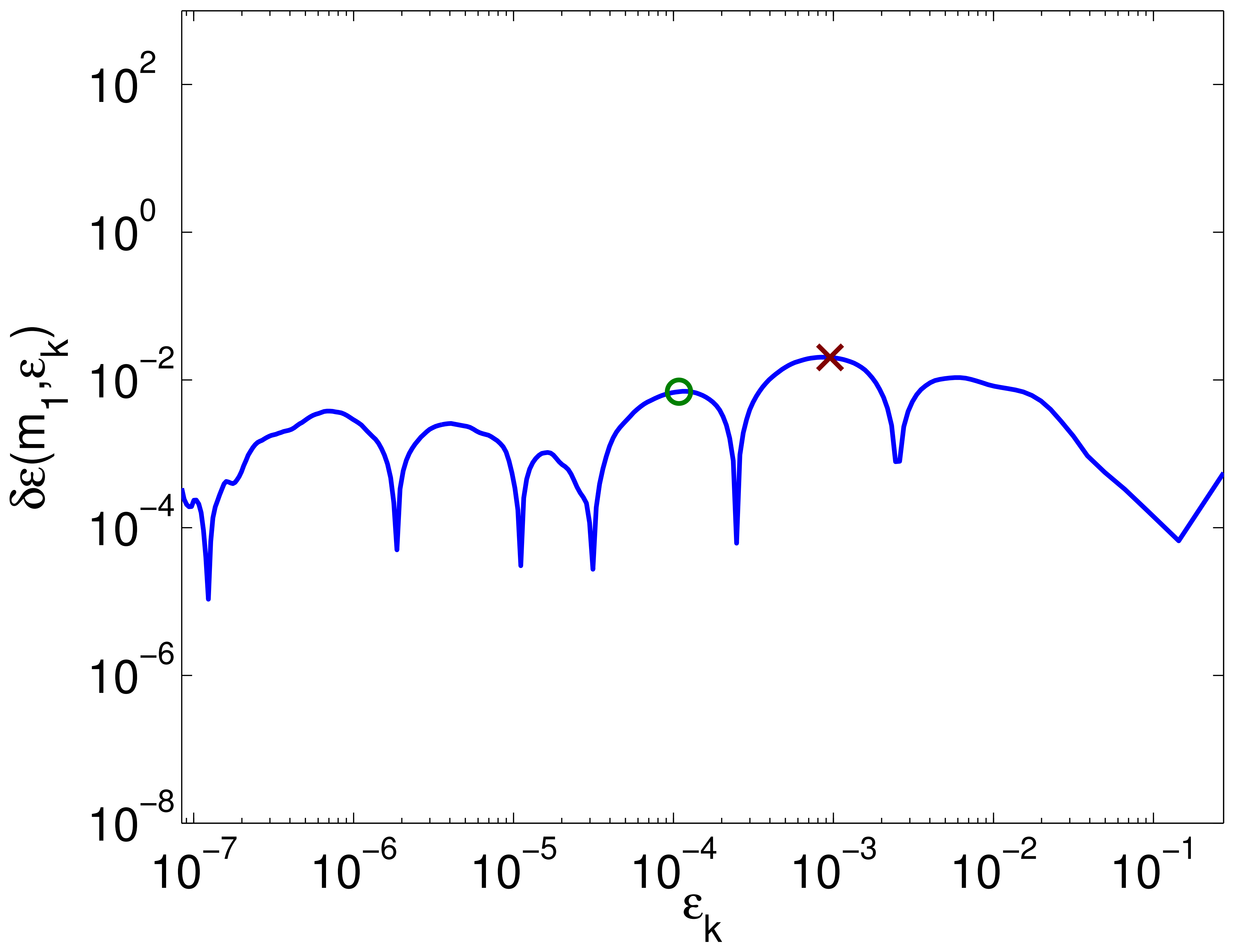
$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

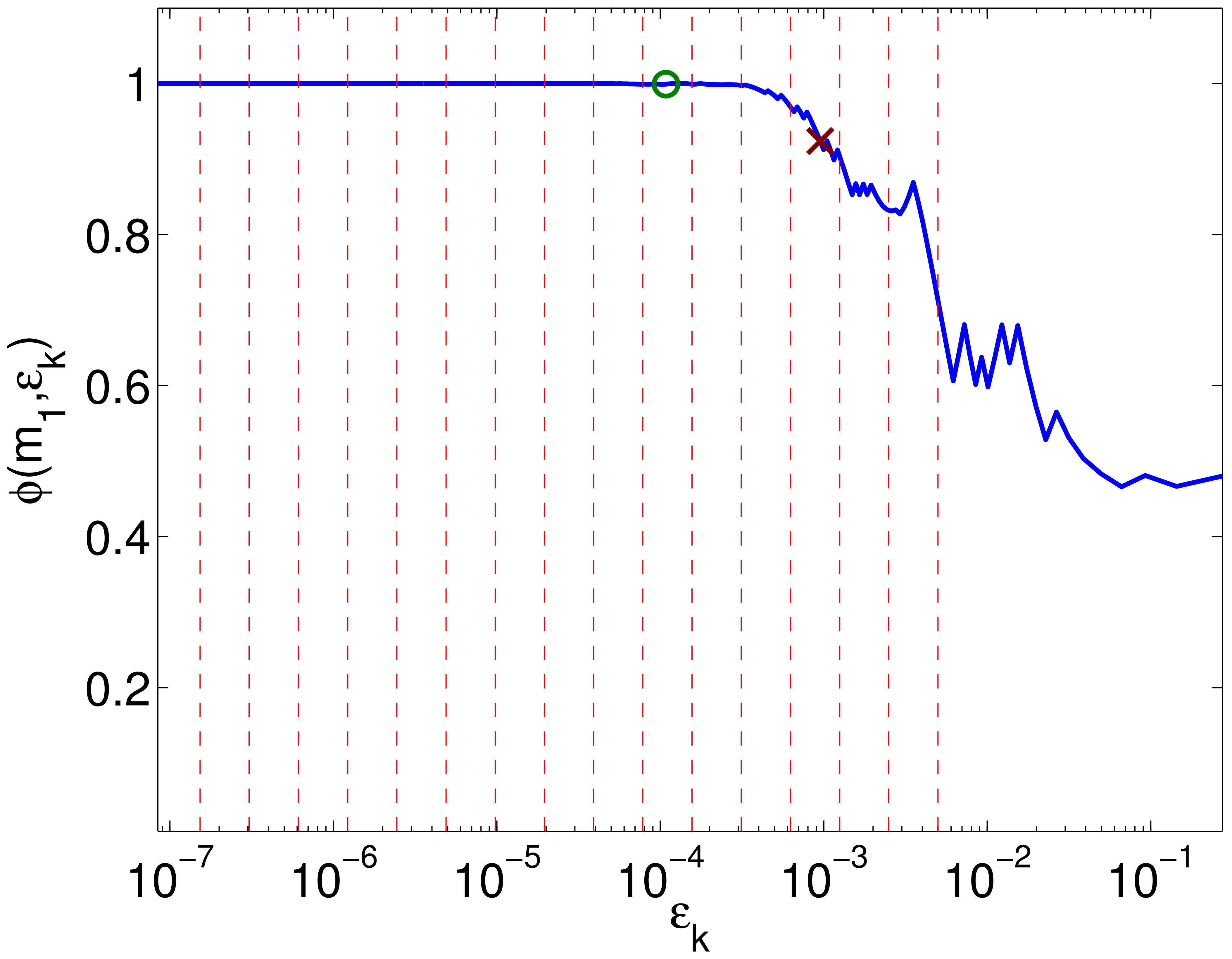


First Order - The $\delta\varepsilon_k$ Criterion

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

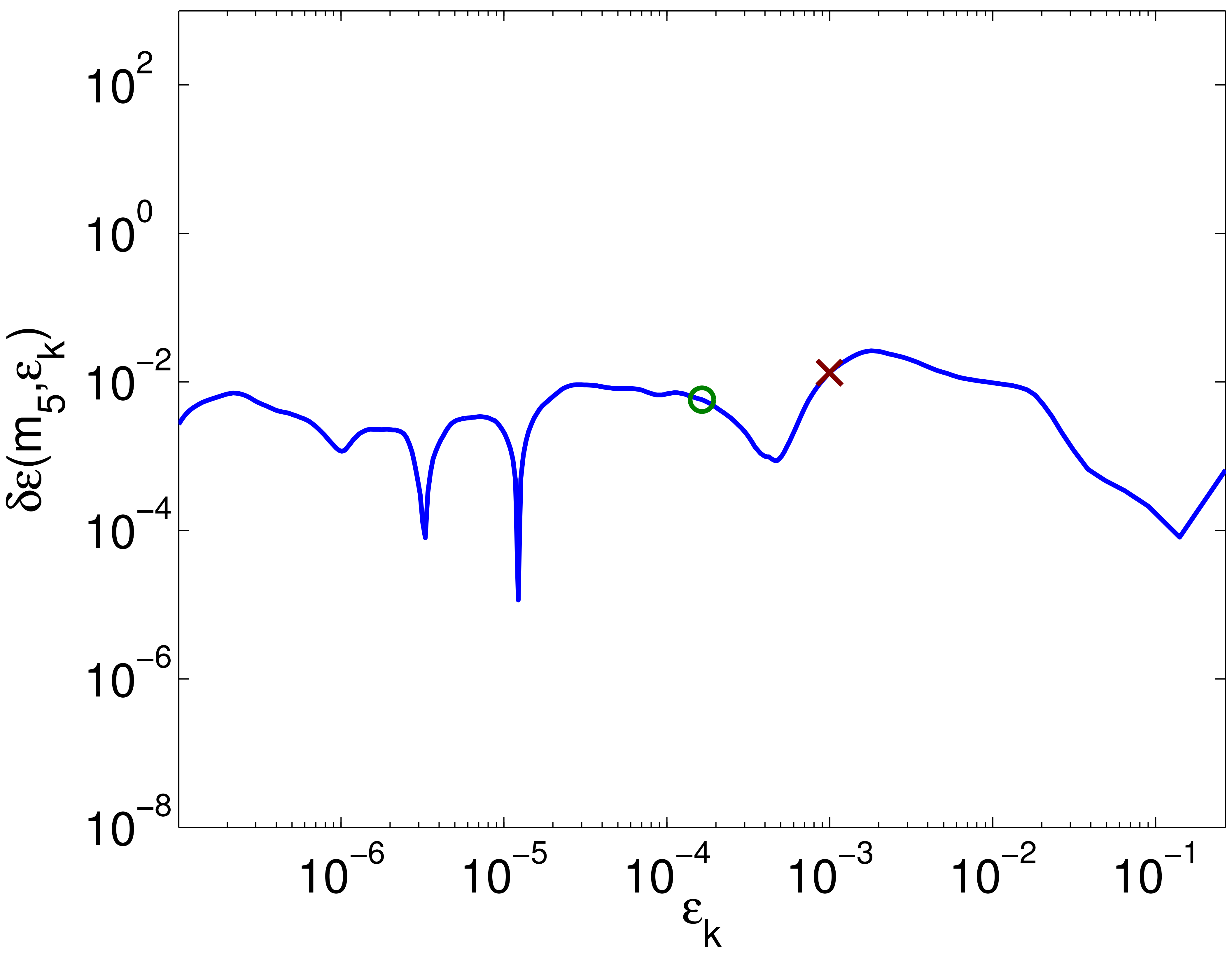


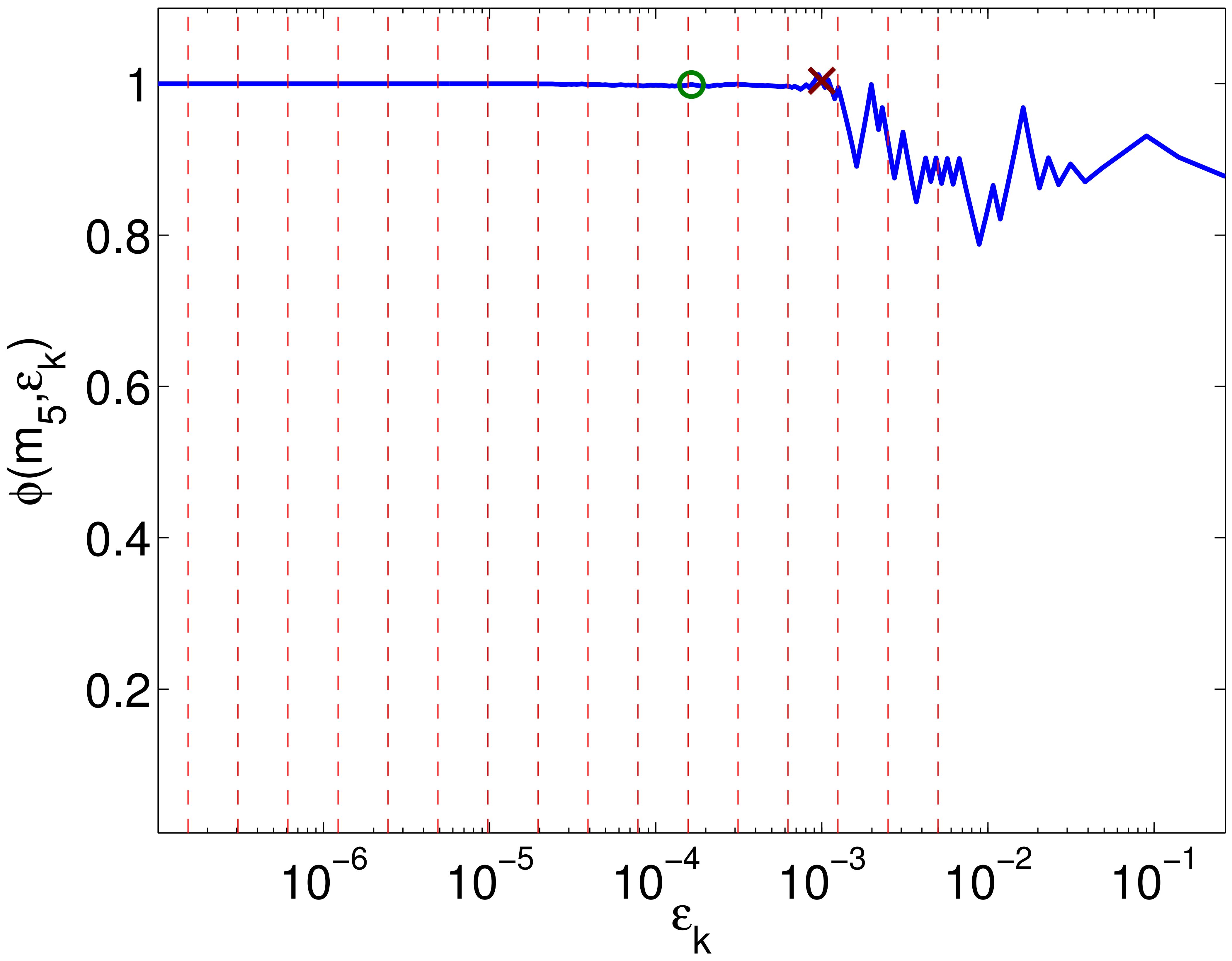


First Order - The $\delta\varepsilon_k$ Criterion

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

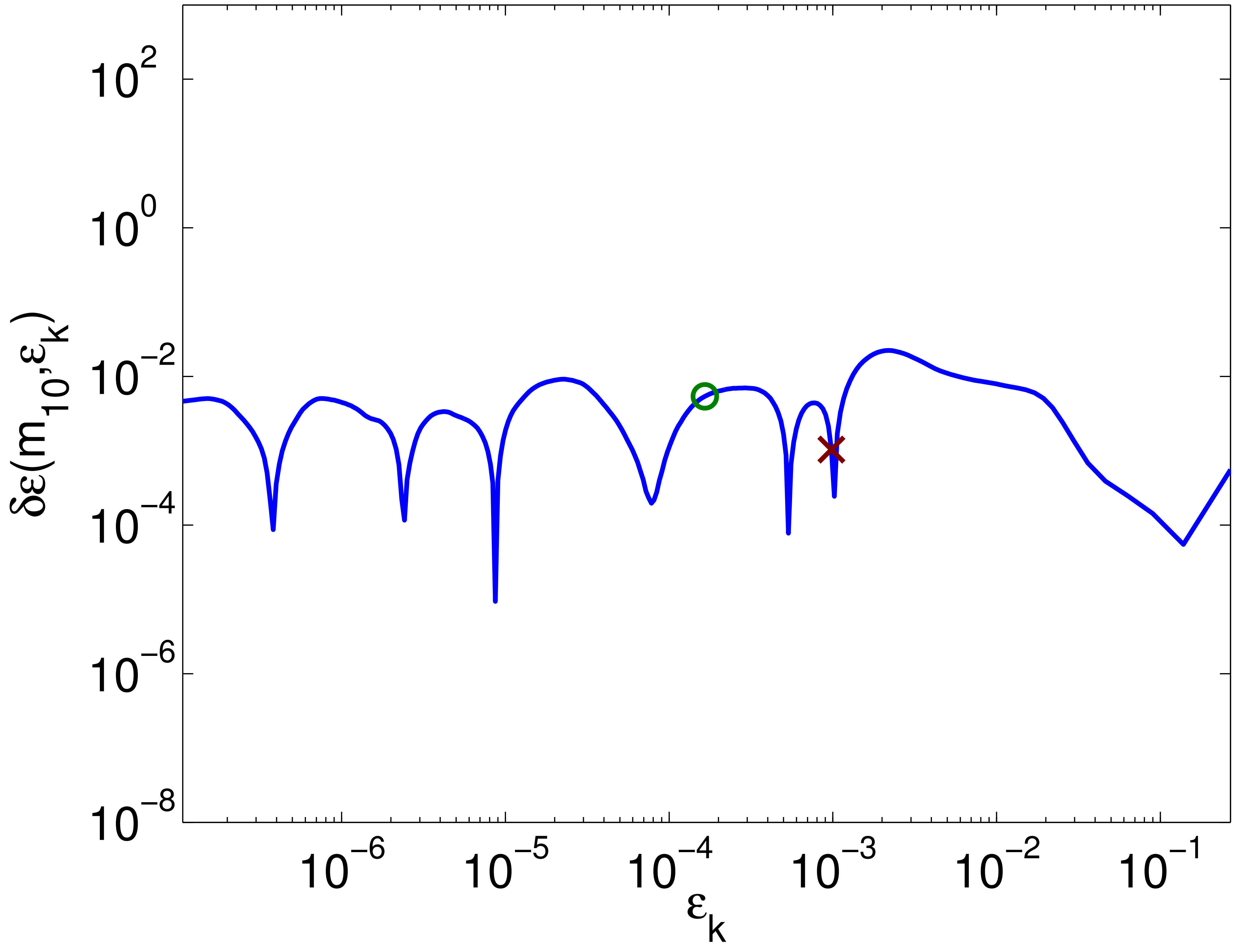
$$\varepsilon_k := \frac{\left\|A(m)u^i - q^i\right\|_2}{\left\|q^i\right\|_2}$$

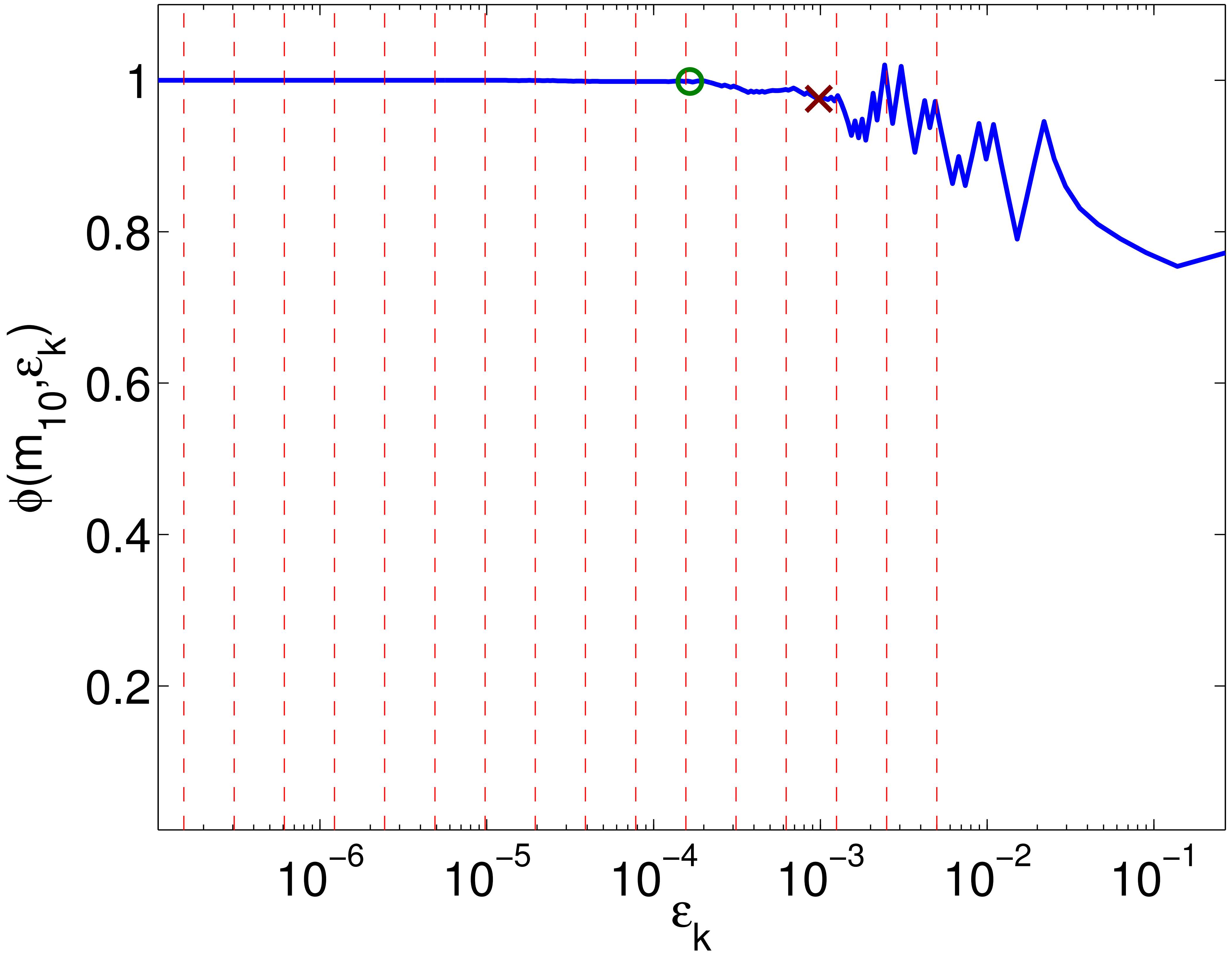




$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

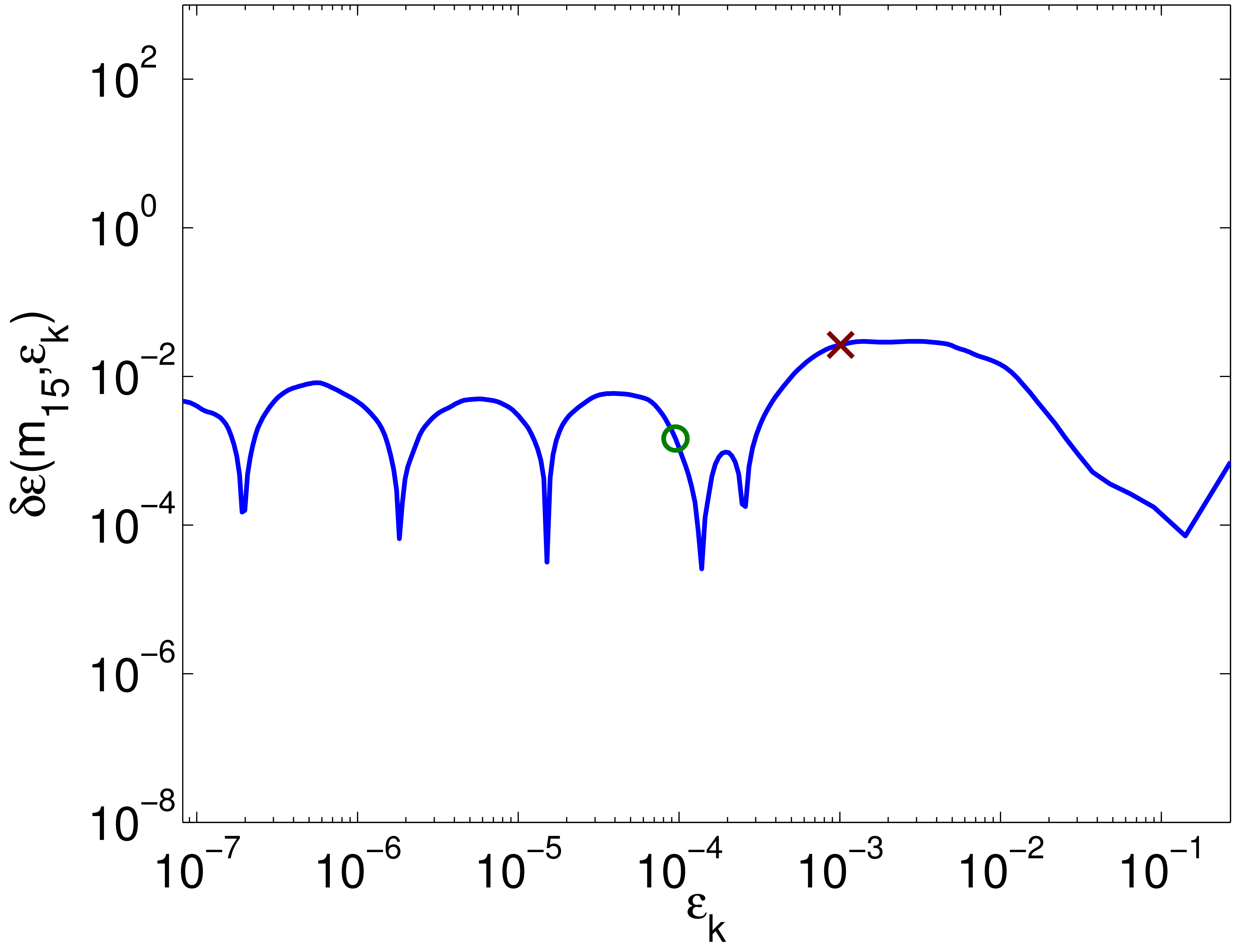


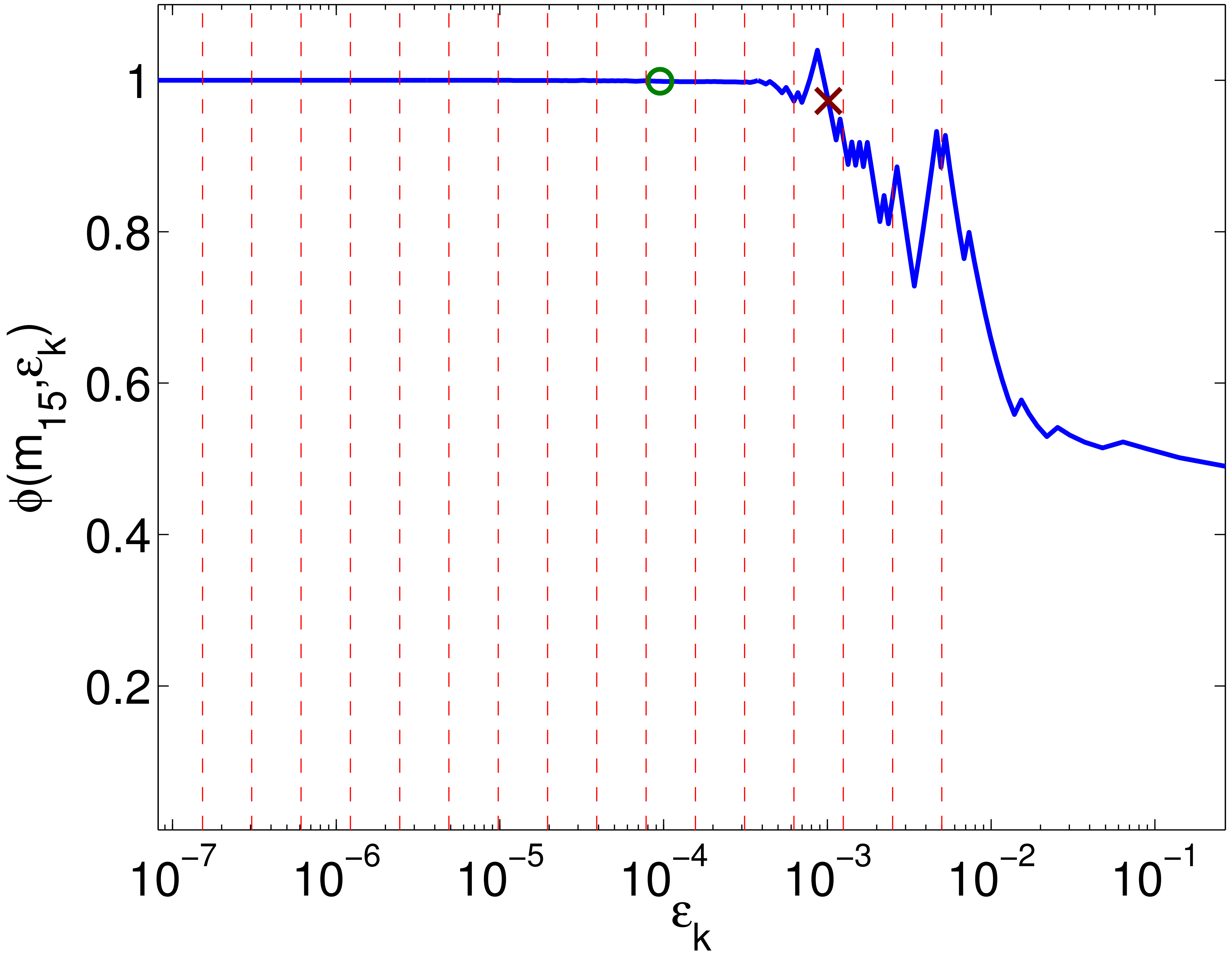


First Order - The $\delta\varepsilon_k$ Criterion

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

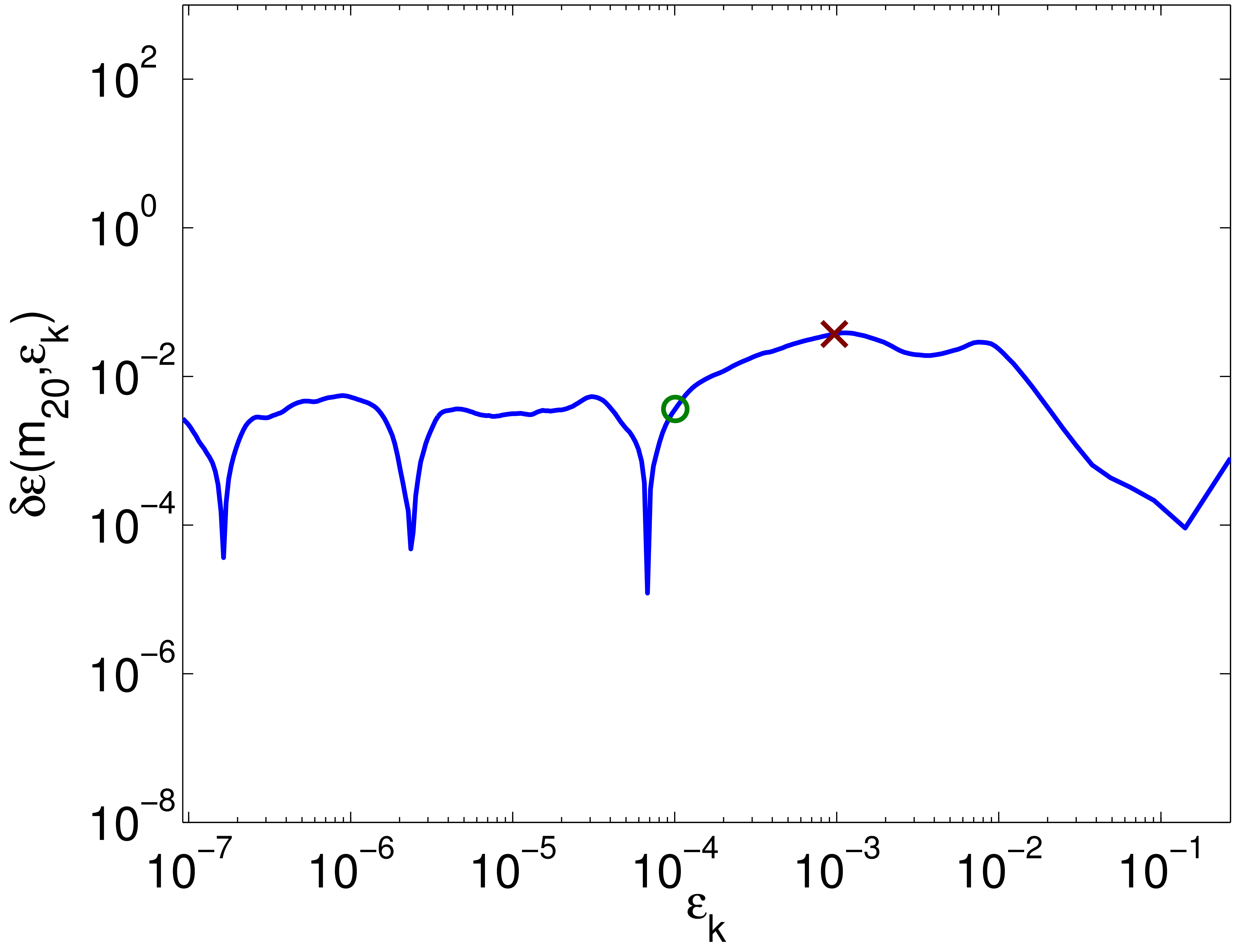


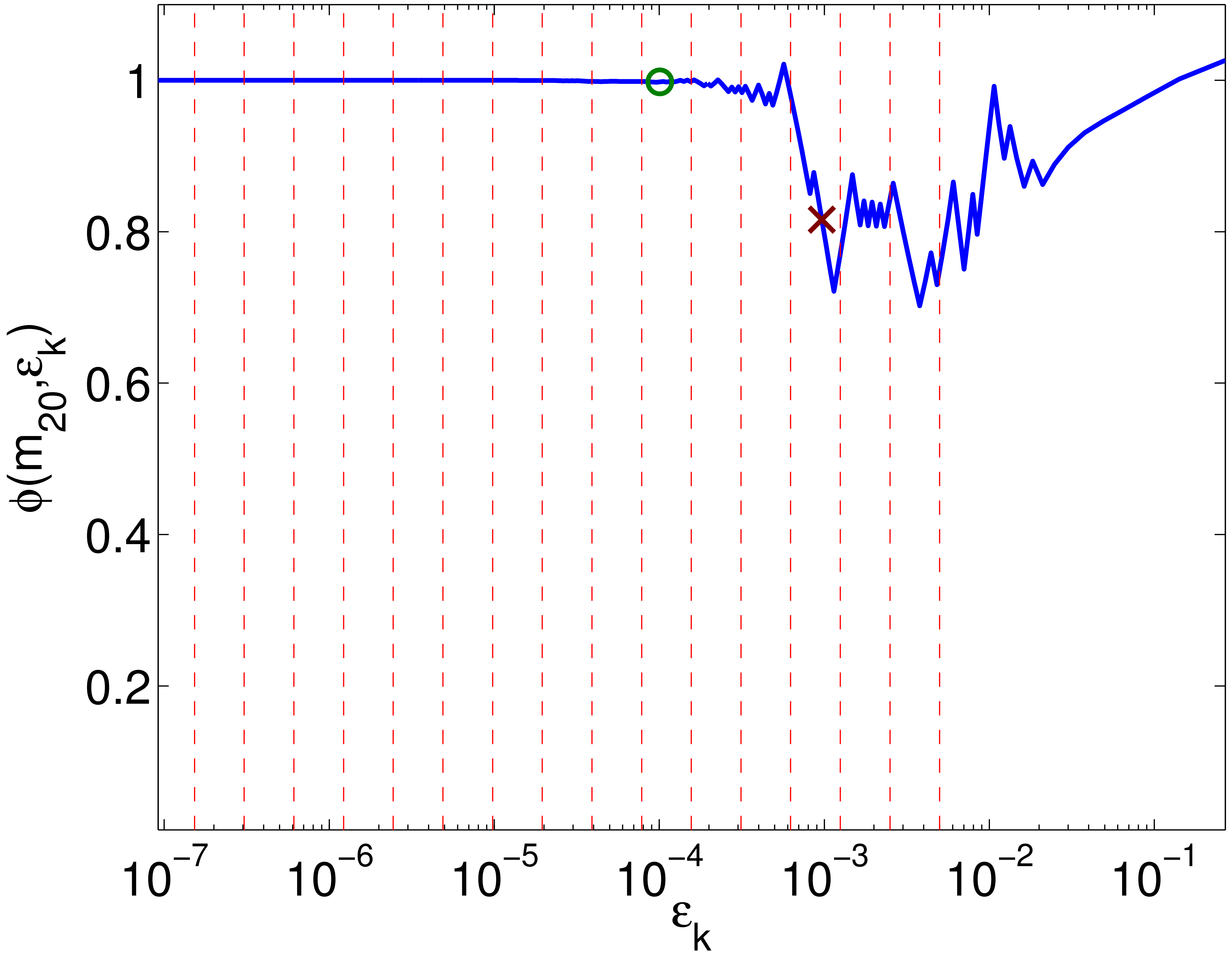


First Order - The $\delta\varepsilon_k$ Criterion

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$

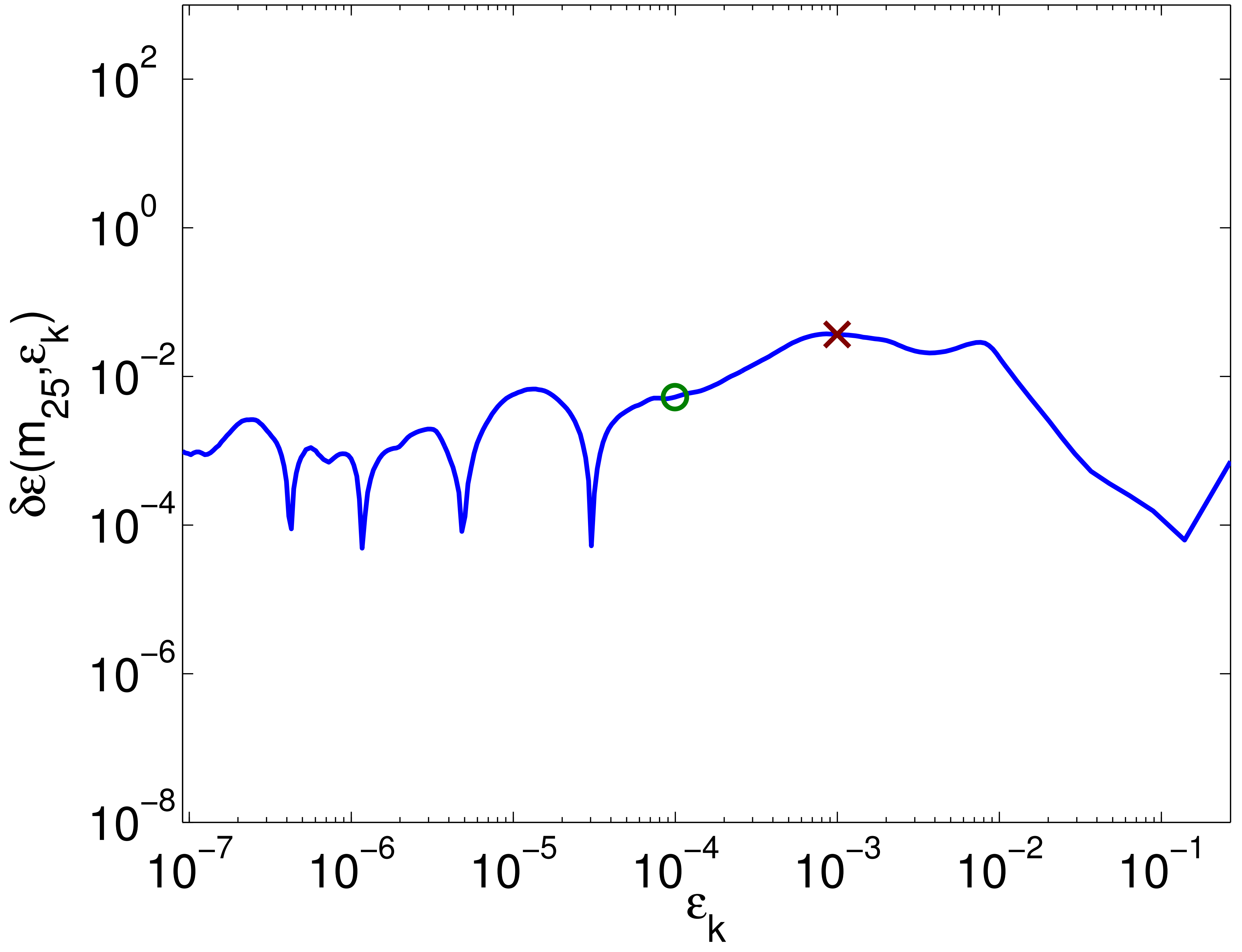


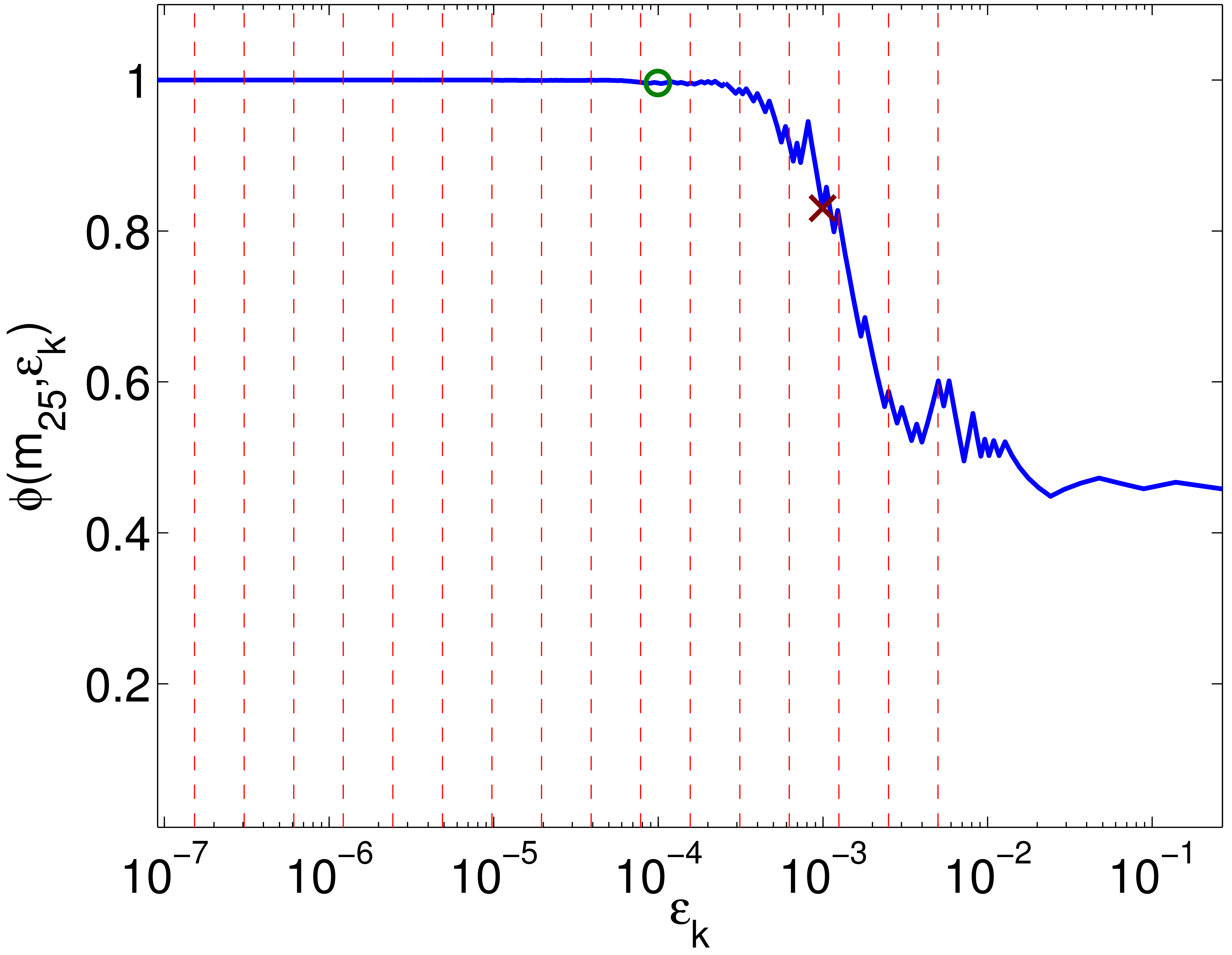


First Order - The $\delta\varepsilon_k$ Criterion

$$\frac{|\Phi(m,\varepsilon_k) - \Phi(m,\varepsilon_{k+1})|}{|\varepsilon_k - \varepsilon_{k+1}|} \leq \eta$$

$$\varepsilon_k := \frac{\left\| A(m)u^i - q^i \right\|_2}{\left\| q^i \right\|_2}$$





Blaming the Model m - The δs_k Criterion

Let

$$A(m)u \approx q \quad \text{s.t.} \quad \frac{\|A(m)u - q\|_2}{\|q\|_2} < \varepsilon$$

however, provided that no entry in u is zero, there is always a matrix E such that

$$(A(m) + E)u = q$$

for instance, defining $r = A(m)u - q$,

$$E = diag\left(\frac{r_i}{u_i}\right).$$

Blaming the Model m - The δs_k Criterion

If there is a perturbation δs such that

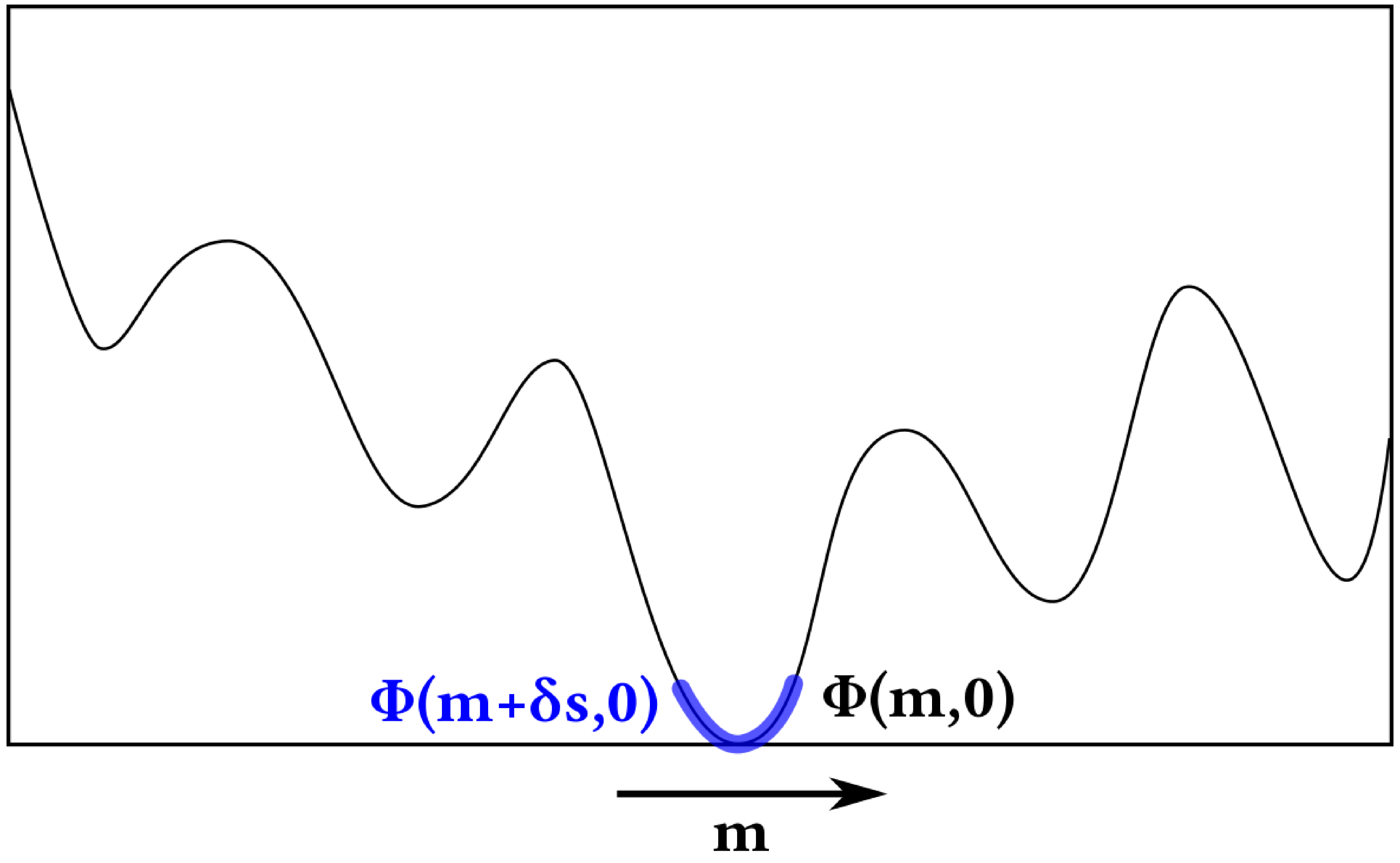
$$A(m + \delta s) = A(m) + E$$

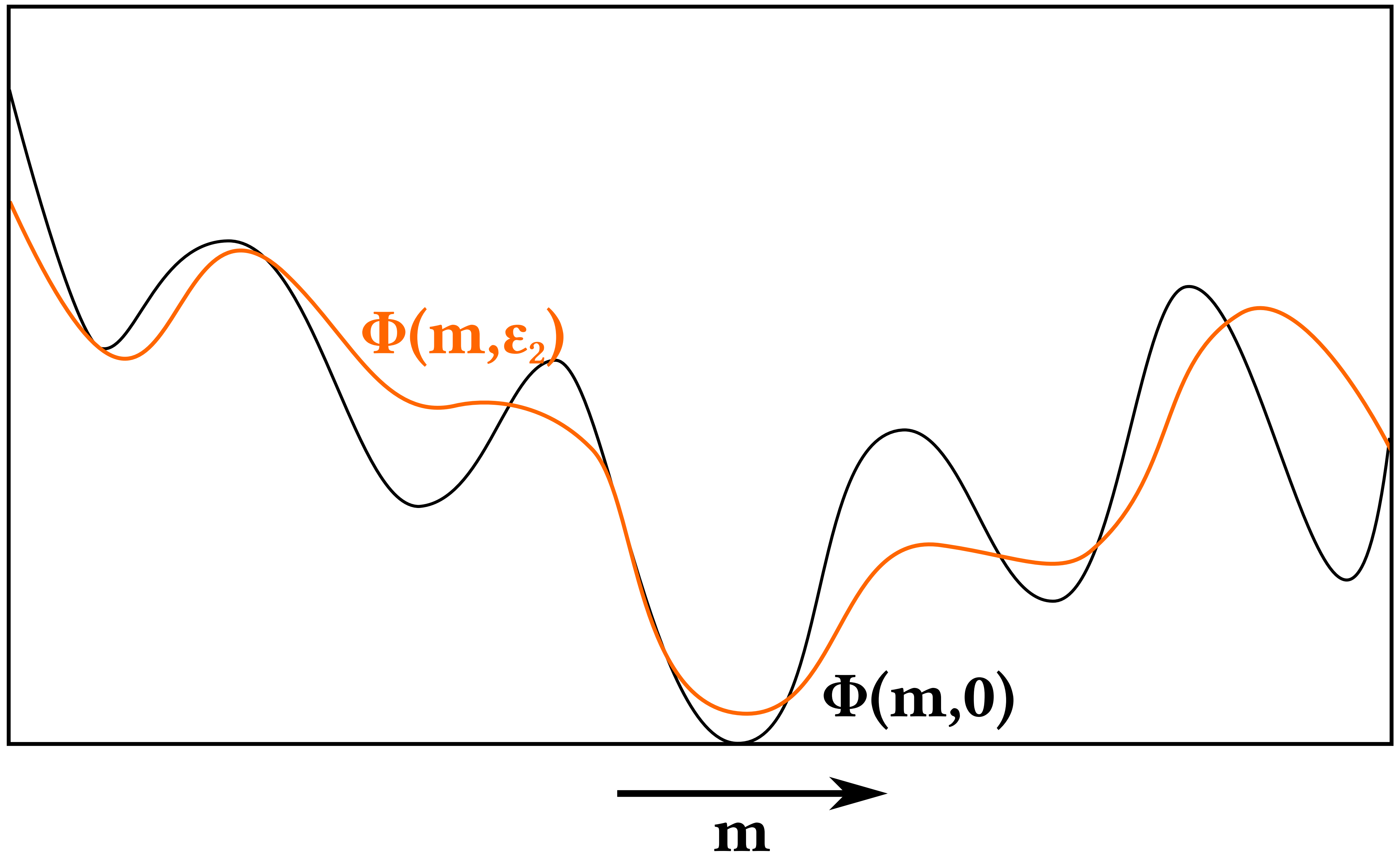
then u is the **exact solution** the model $m + \delta s$, that is

$$\Phi(m + \delta s, 0) = \Phi(m, \varepsilon)$$

We thus want to choose ε such that **δs is small**.
Finding **δs** (or an approximation) may be possible depending on the discretization scheme.

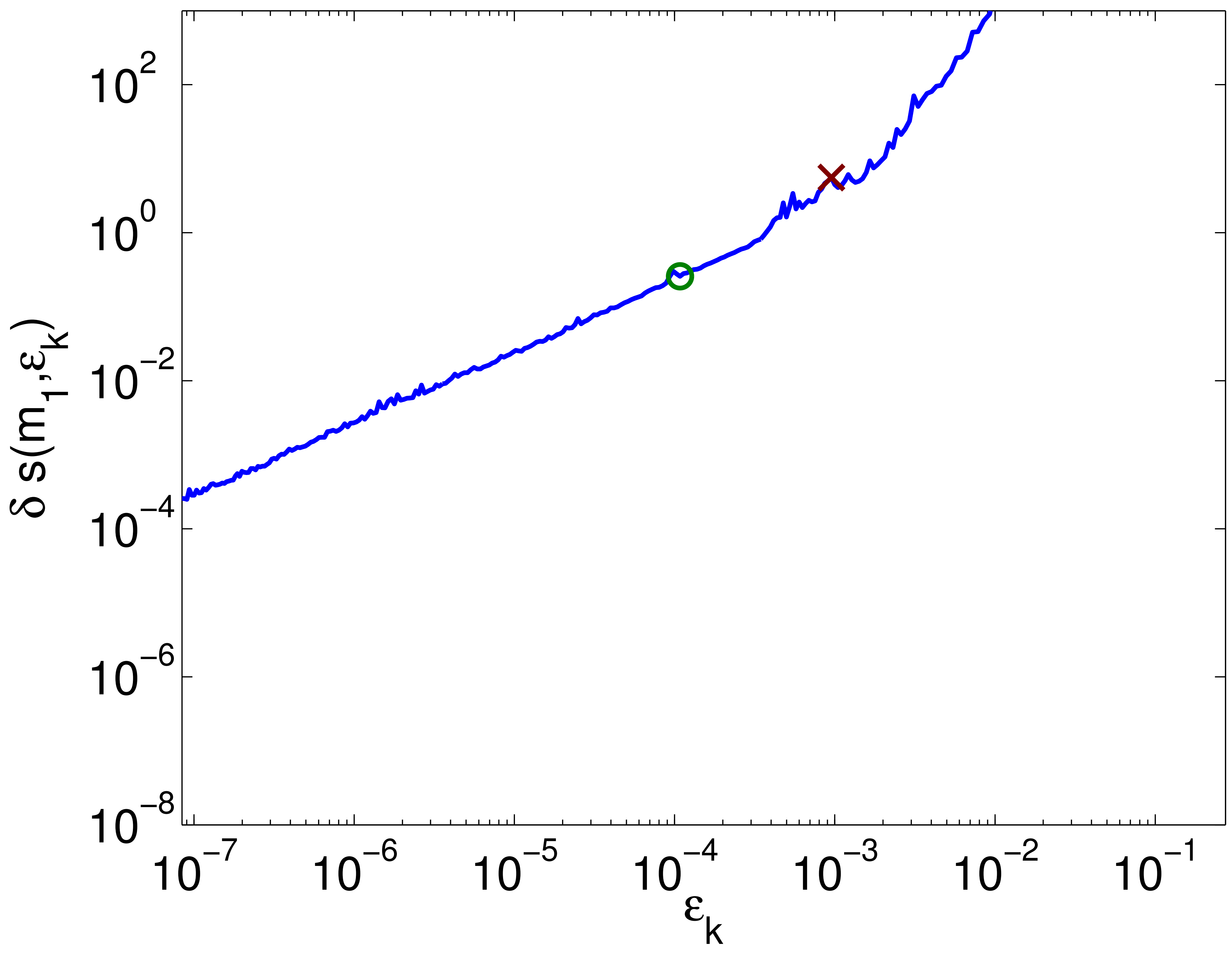
Blaming the Model m - The δs_k Criterion

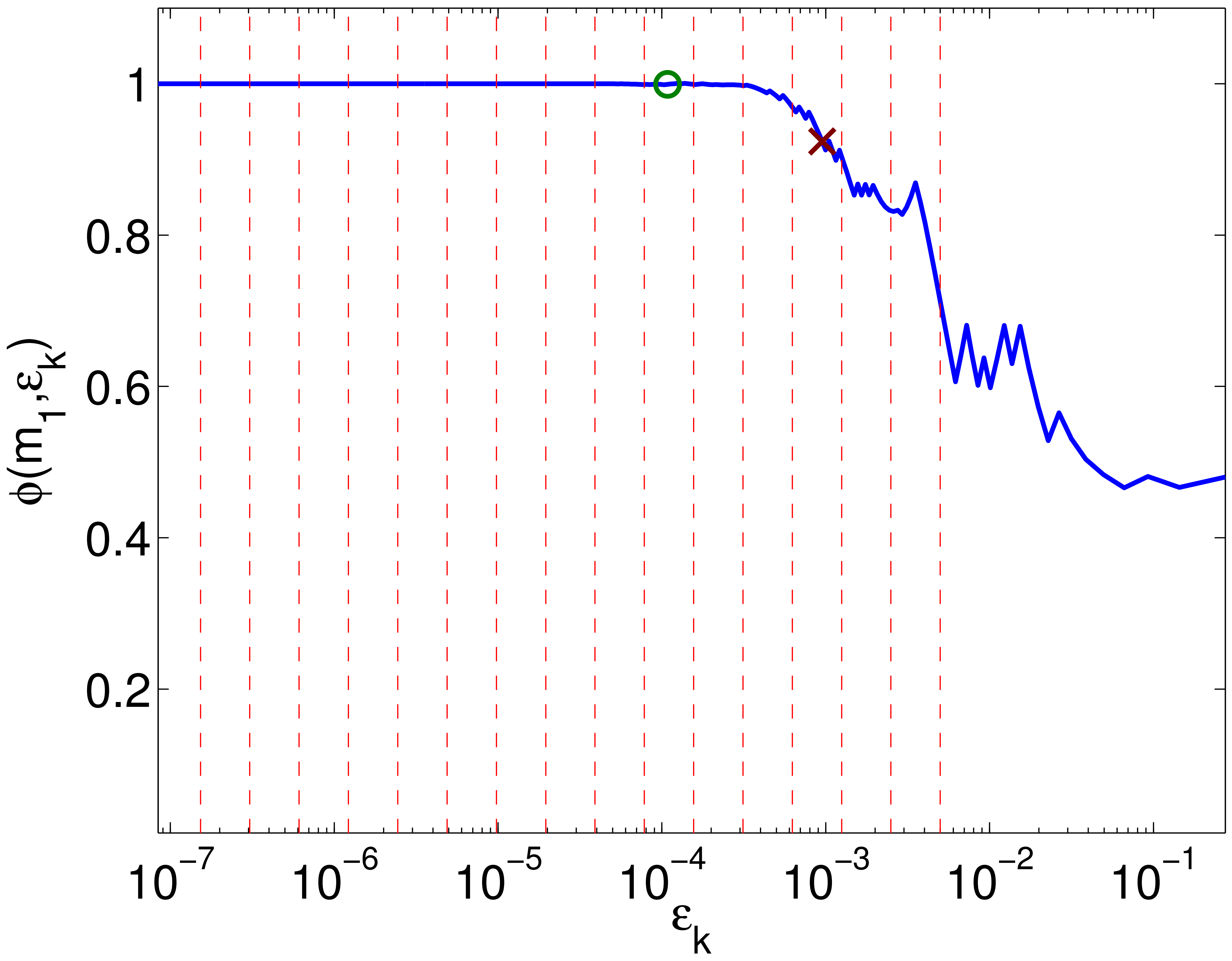




$$\left|\frac{\delta s_k}{m}\right| \leq \eta \qquad \delta s_k := \frac{1}{C\omega^2}\frac{A(m)u^i - q^i}{u^i}$$

(pointwise division)

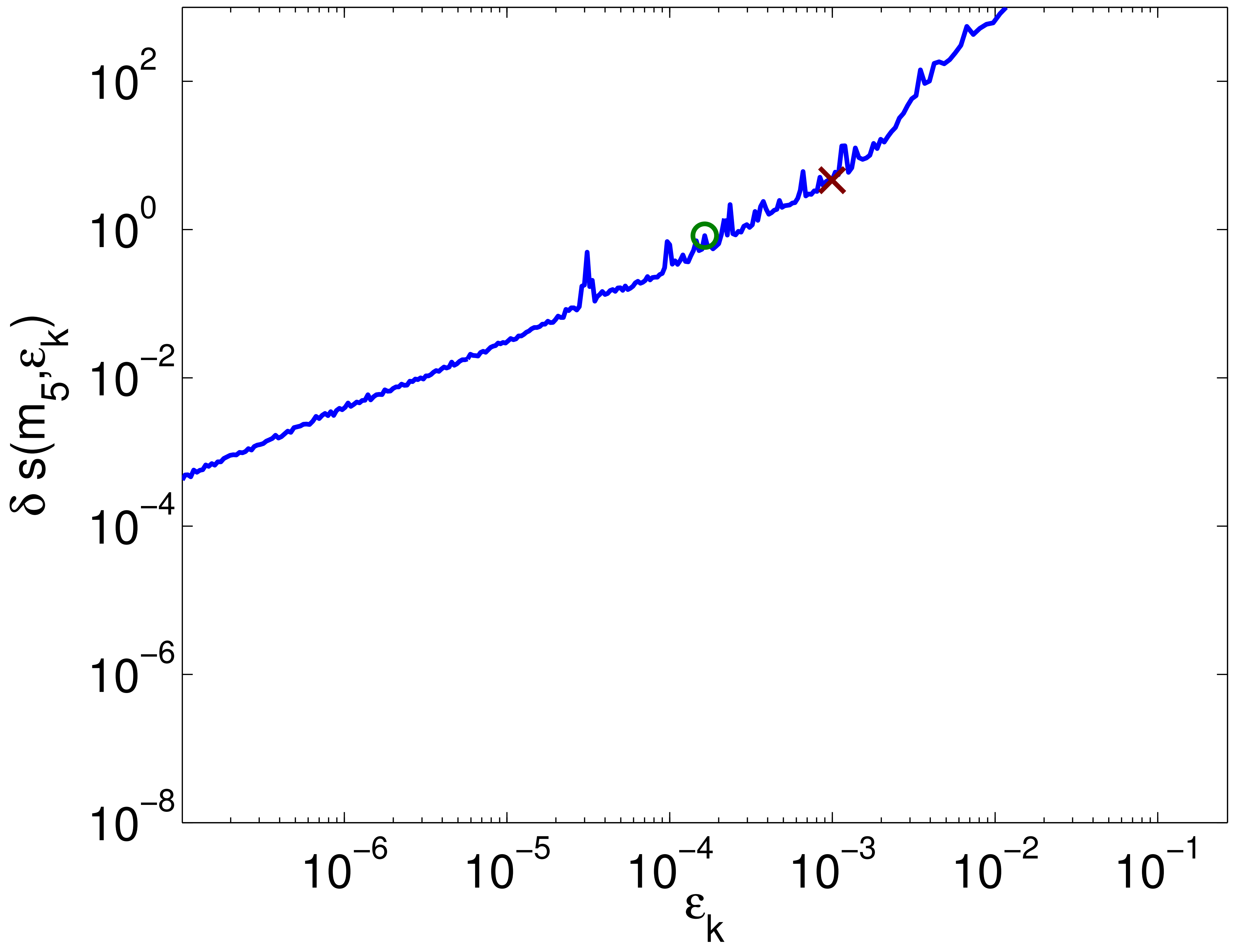


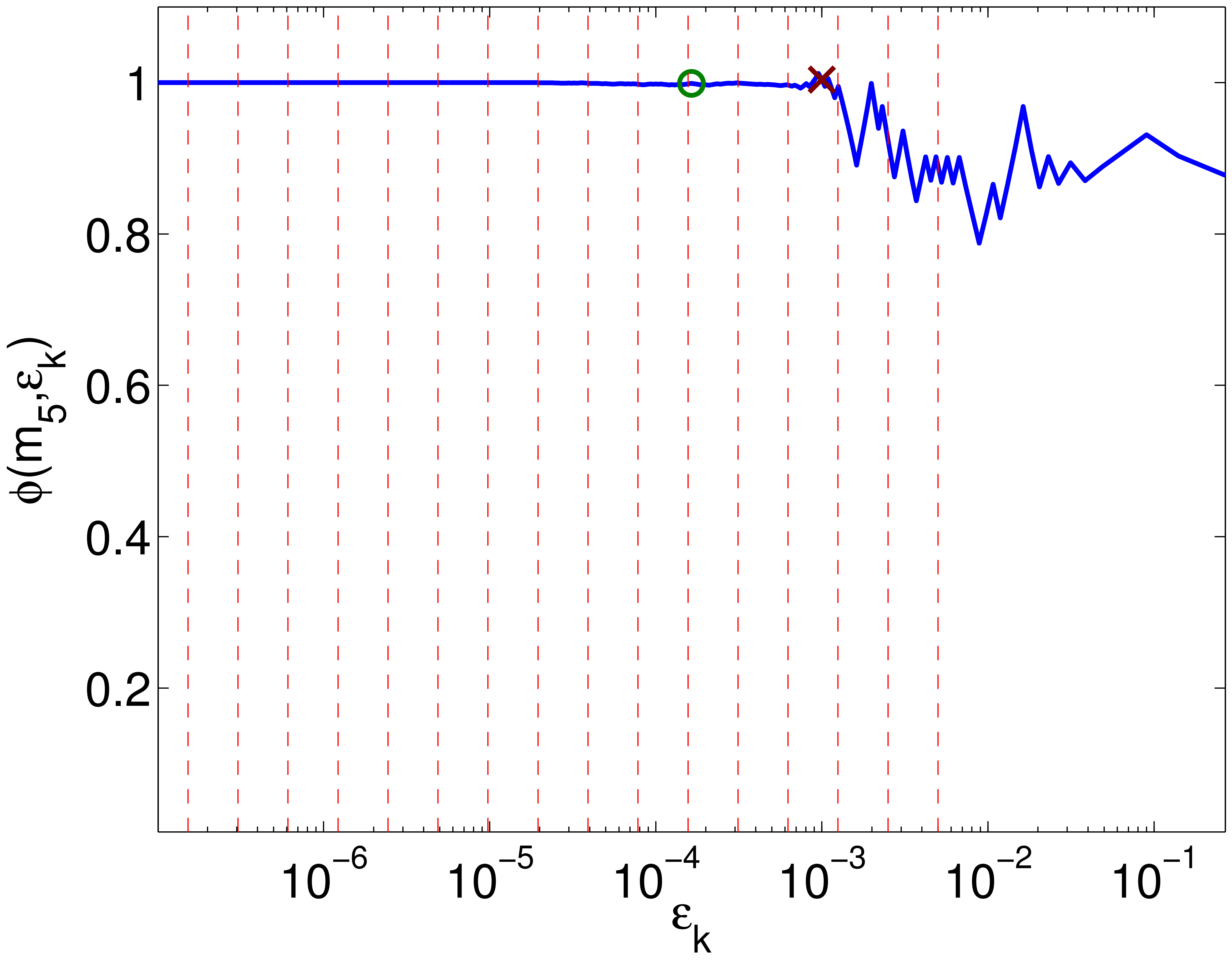


Blaming the Model m - The δs_k Criterion

$$\left|\frac{\delta s_k}{m}\right| \leq \eta \qquad \delta s_k := \frac{1}{C\omega^2}\frac{A(m)u^i - q^i}{u^i}$$

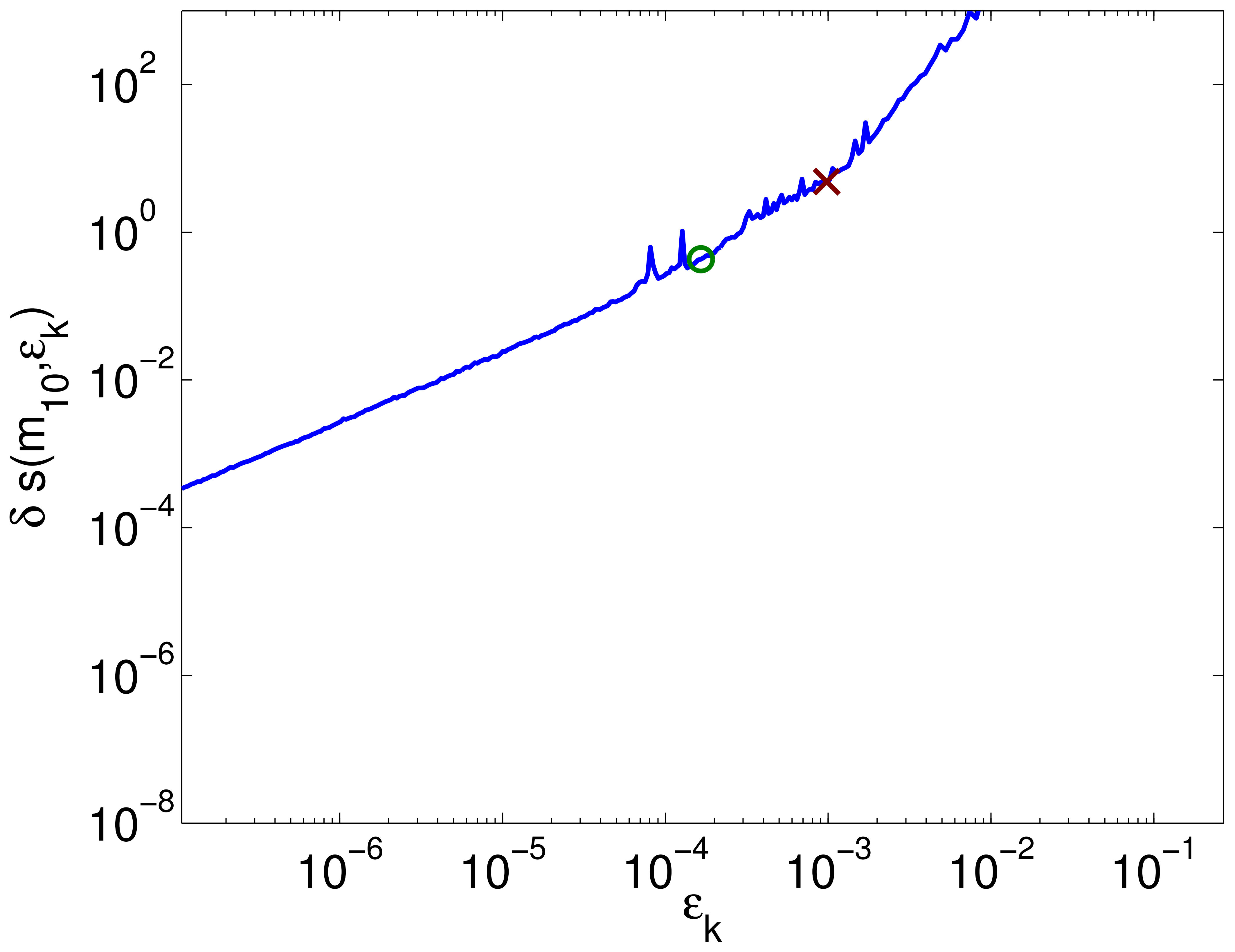
(pointwise division)

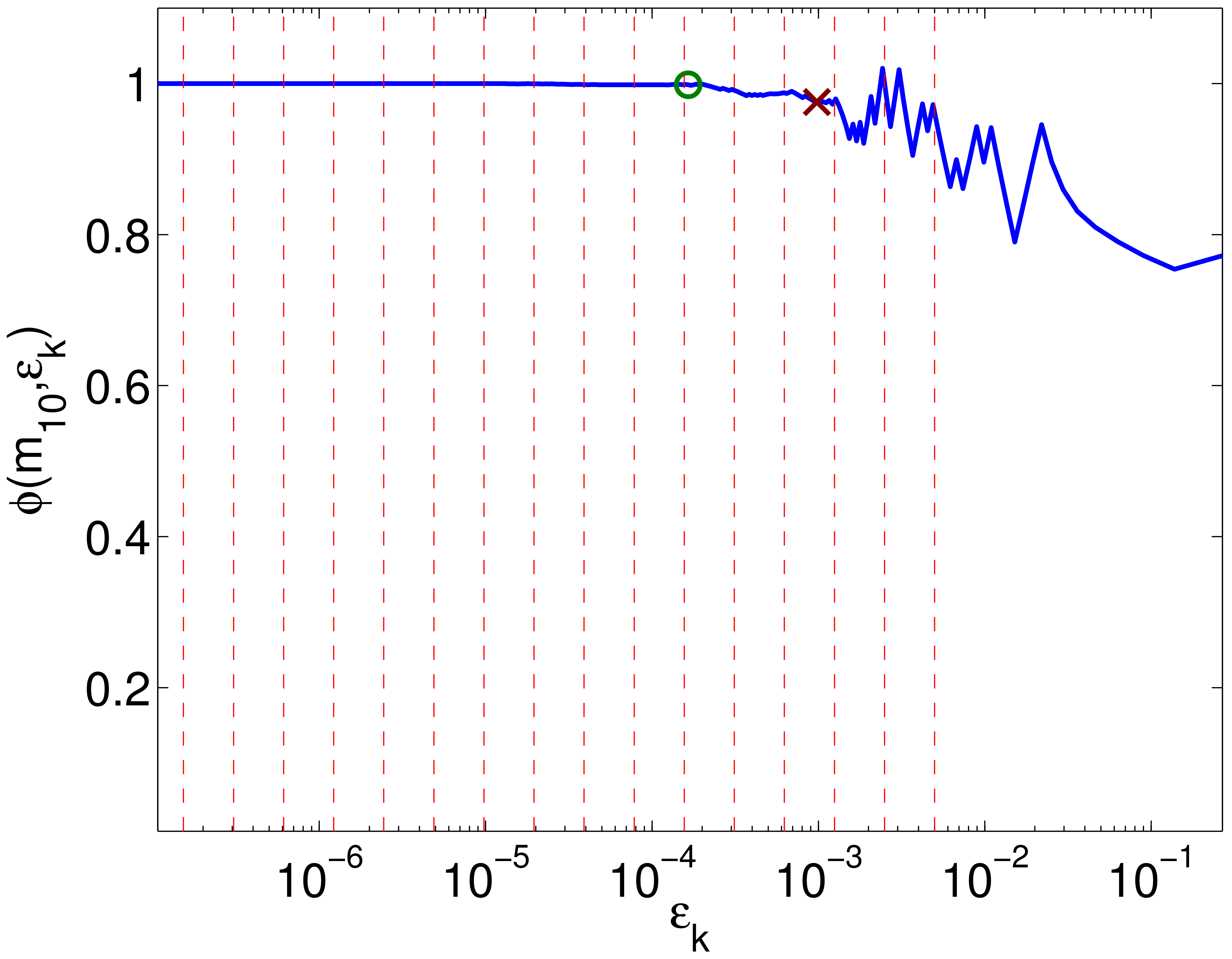




$$\left|\frac{\delta s_k}{m}\right| \leq \eta \qquad \delta s_k := \frac{1}{C\omega^2}\frac{A(m)u^i - q^i}{u^i}$$

(pointwise division)

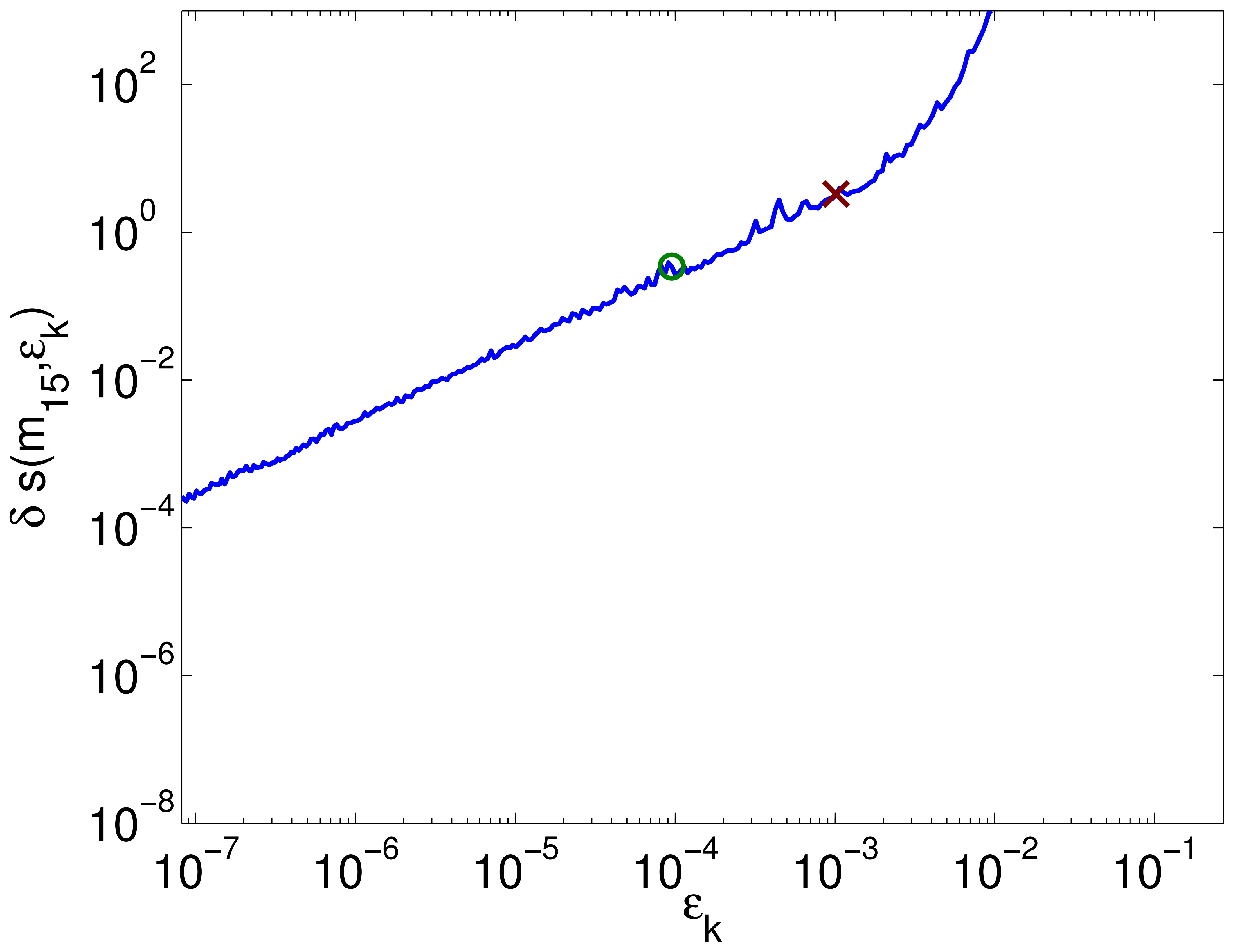


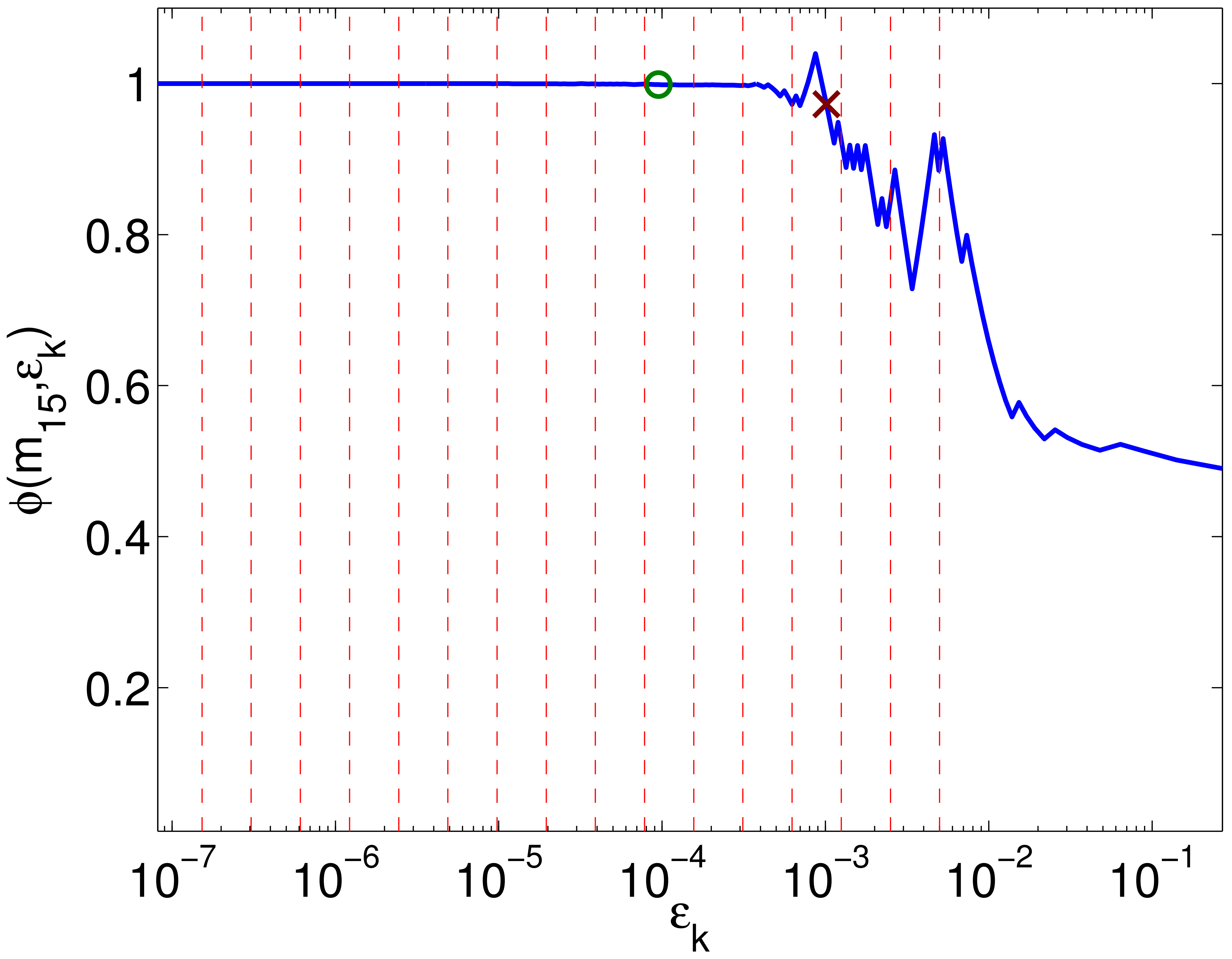


Blaming the Model m - The δs_k Criterion

$$\left|\frac{\delta s_k}{m}\right| \leq \eta \qquad \delta s_k := \frac{1}{C\omega^2}\frac{A(m)u^i - q^i}{u^i}$$

(pointwise division)

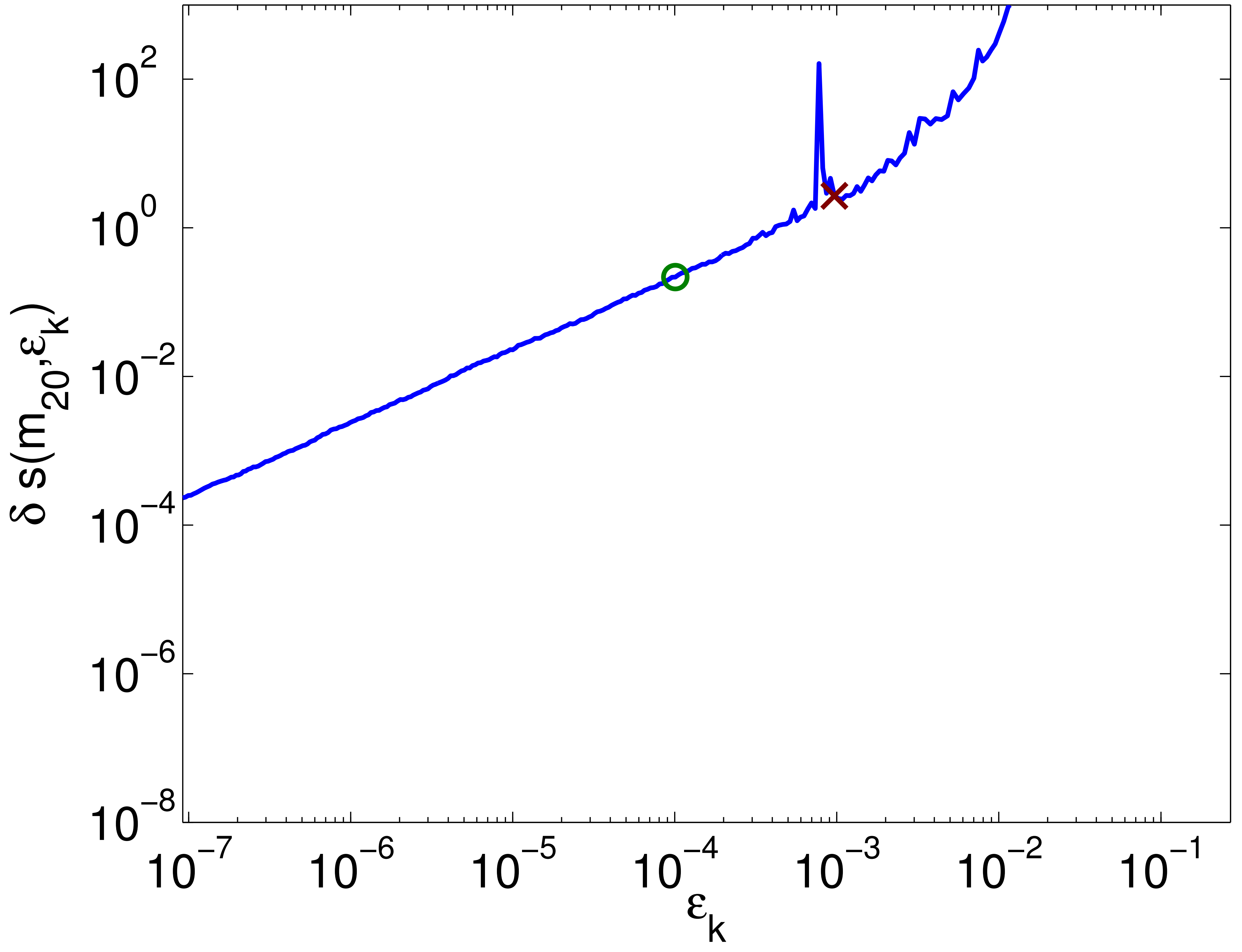


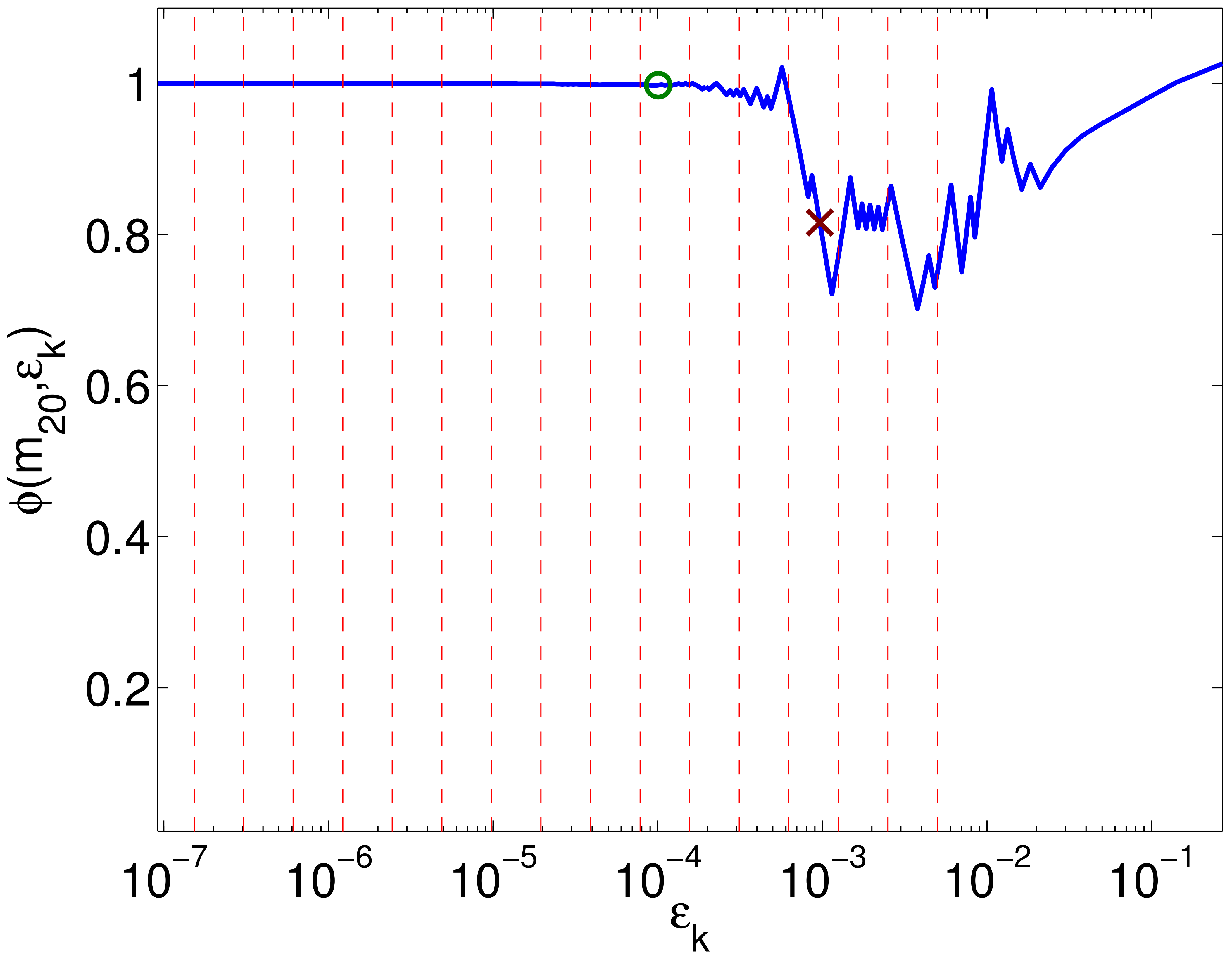


Blaming the Model m - The δs_k Criterion

$$\left|\frac{\delta s_k}{m}\right| \leq \eta \qquad \delta s_k := \frac{1}{C\omega^2}\frac{A(m)u^i - q^i}{u^i}$$

(pointwise division)

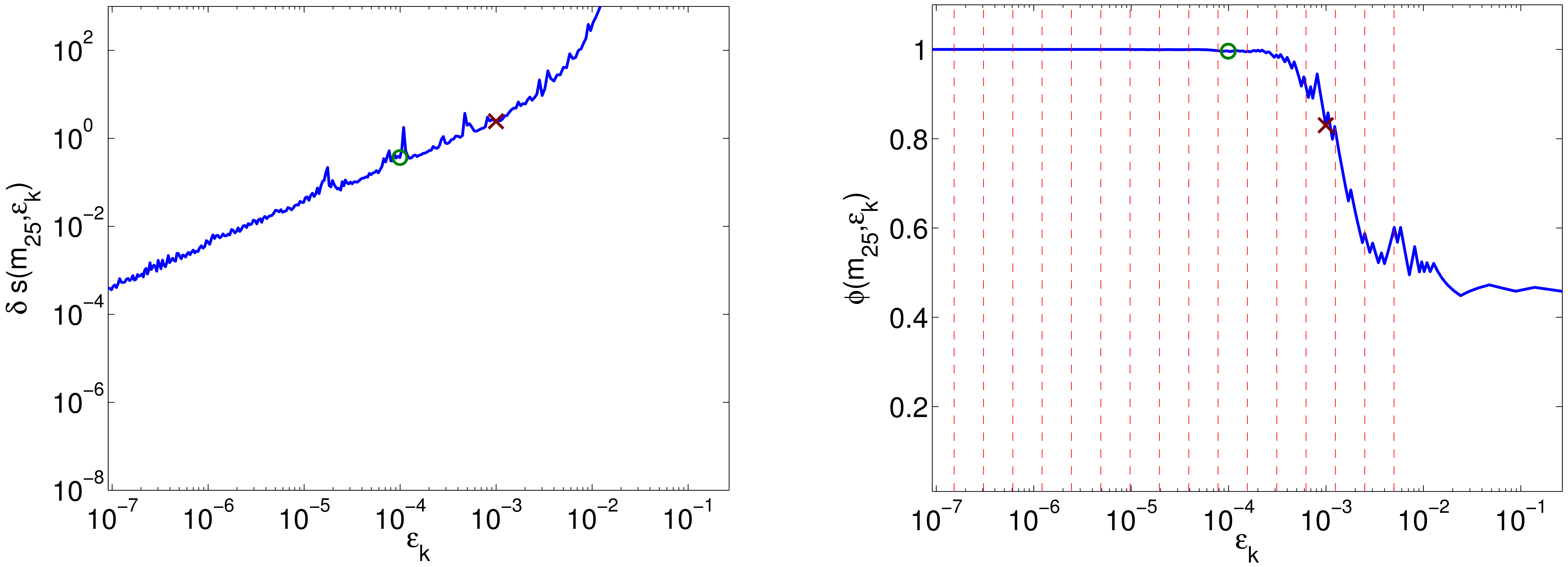




Blaming the Model m - The δs_k Criterion

$$\left|\frac{\delta s_k}{m}\right| \leq \eta \qquad \delta s_k := \frac{1}{C\omega^2}\frac{A(m)u^i - q^i}{u^i}$$

(pointwise division)



Conclusions

Questions?

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Total SA, WesternGeco, and Woodside.