OPTIMIZATION AND RICHARDSON EXTRAPOLATION - Cont'd.

- **Ex.:** Richardson extrapolation is simple linear extrapolation when only two polyline tangent curve approximations are used! Consider $\vec{u}(x,y) = \begin{pmatrix} 2y \\ -2x \end{pmatrix}$ as gradient field.

  - i) Polyline 1, stepsize Δ, 3 points:

$$\mathbf{x}_0^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_1^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Delta \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2\Delta \\ 1 \end{pmatrix},$$

$$\mathbf{x}_2^1 = \begin{pmatrix} 2\Delta \\ 1 \end{pmatrix} + \Delta \begin{pmatrix} 2 \\ -4\Delta \end{pmatrix} = \begin{pmatrix} 4\Delta \\ 1-4\Delta^2 \end{pmatrix}$$

$$\overset{\Delta = 1/2}{\Longrightarrow} \quad \underline{\underline{\mathbf{x}_0^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_1^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2^1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}}}$$

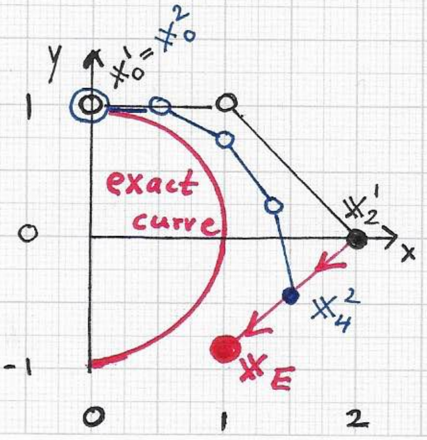
  - ii) Polyline 2, stepsize $\Delta/2$, 5 points:

$$\mathbf{x}_0^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_1^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{\Delta}{2} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta \\ 1 \end{pmatrix},$$

$$\mathbf{x}_2^2 = \ldots = \begin{pmatrix} 2\Delta \\ 1-\Delta^2 \end{pmatrix}, \quad \mathbf{x}_3^2 = \begin{pmatrix} \Delta(3-\Delta^2) \\ 1-3\Delta^2 \end{pmatrix}$$

$$\mathbf{x}_4^2 = \ldots = \begin{pmatrix} 4\Delta(1-\Delta^2) \\ 1-6\Delta^2+\Delta^4 \end{pmatrix}$$

$$\overset{\Delta = 1/2}{\Longrightarrow} \quad \underline{\mathbf{x}_0^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_1^2 = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}, \quad \mathbf{x}_2^2 = \begin{pmatrix} 1 \\ 3/4 \end{pmatrix}, \quad \mathbf{x}_3^2 = \begin{pmatrix} 11/8 \\ 1/4 \end{pmatrix}, \quad \mathbf{x}_4^2 = \begin{pmatrix} 3/2 \\ -7/16 \end{pmatrix}}$$



$\Rightarrow$ Richardson extrapolation = linear extrapolation:

$$\underline{\underline{\mathbf{x}_E}} = 2 \cdot \mathbf{x}_4^2 - 1 \cdot \mathbf{x}_2^1 = \begin{pmatrix} 3 \\ -7/8 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \underline{\underline{\begin{pmatrix} 1 \\ -7/8 \end{pmatrix}}} = \bullet$$

■ OPTIMIZATION AND RICHARDSON EXTRAPOLATION – Cont'd.

• Ex.: "Pathological case" – Tangent curves of gradient vector field of $f(x,y) = x^2+y^2$, i.e., $\vec{u} = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$, with (exact) minimum at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ (start: $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$) $\Rightarrow$ steepest descent: $-\vec{u} = \begin{pmatrix} -2x \\ -2y \end{pmatrix}$

i) Polyline 0, 2 pnts, Δ: $x_0^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $x_1^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \Delta \begin{pmatrix} -2 \\ -2 \end{pmatrix} = \begin{pmatrix} 1-2\Delta \\ " \end{pmatrix}$

ii) " 1, 3 pnts, $\frac{\Delta}{2}$: $x_0^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ldots, x_2^1 = \begin{pmatrix} (1-\Delta)^2 \\ " \end{pmatrix}$

iii) " 2, 5 ", $\frac{\Delta}{4}$: $x_0^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ldots, x_4^2 = \begin{pmatrix} (1-\Delta/2)^4 \\ " \end{pmatrix}$

iv) " 3, 9 ", $\frac{\Delta}{8}$: $x_0^3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ldots, x_8^3 = \begin{pmatrix} (1-\Delta/4)^8 \\ " \end{pmatrix}$

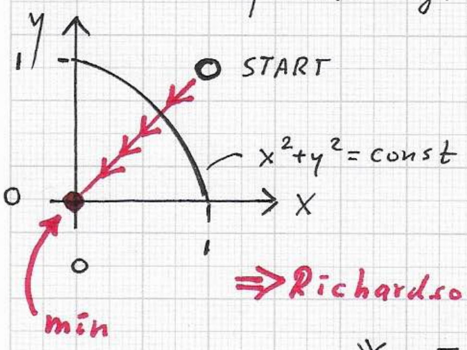
v) " 4, 17 ", $\frac{\Delta}{16}$: $x_0^4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ldots, x_{16}^4 = \begin{pmatrix} (1-\Delta/8)^{16} \\ " \end{pmatrix}$

$$\overset{\Delta=1}{\Longrightarrow} \underline{\underline{x_{2^N}^N = \begin{pmatrix} \left((2^{N-1}-1)/2^{N-1}\right)^{2^N} \\ " \end{pmatrix}}}$$

• specifically (for $\Delta = 1$): $\underline{x_1^0 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}}$, $\underline{x_2^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}}$, $\underline{x_4^2 = \begin{pmatrix} 1/16 \\ 1/16 \end{pmatrix}} = \begin{pmatrix} .0625 \\ .0625 \end{pmatrix}$

$\underline{x_8^3 = \begin{pmatrix} (3/4)^8 \\ " \end{pmatrix}} \approx \begin{pmatrix} .1001 \\ " \end{pmatrix}$

$\underline{x_{16}^4 = \begin{pmatrix} (7/8)^{16} \\ " \end{pmatrix}} \approx \begin{pmatrix} .1181 \\ .1181 \end{pmatrix}$



$\Rightarrow$ Richardson extrapolation:

$$\underline{\underline{x_E}} = \left(1024\, x_{16}^4 - 960\, x_8^3 + 280\, x_4^2 - 30\, x_2^1 + 1 \cdot x_1^0\right) / 315$$

$$\approx \underline{\underline{\begin{pmatrix} .0755 \\ " \end{pmatrix}}} \longrightarrow \text{slow convergence!}$$

$\Rightarrow$ Compare x_E with $\underline{\underline{x_{2^{11}}^{11}}} = \begin{pmatrix} (1023/1024)^{2048} \\ " \end{pmatrix} \approx \underline{\underline{\begin{pmatrix} .1352 \\ " \end{pmatrix}}}$

$\Rightarrow$ Extrapolation still advantageous!

■ OPTIMIZATION AND RICHARDSON EXTRAPOLATION – Cont'd.

- Ex.: "Interplay between gradient vector field and stepsize" –
  – 'Getting close' to a loss function minimum can be difficult and thus time-consuming

  $f(x,y) = x^2 + y^2$, $\text{grad } f = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$, $\vec{u} = -\text{grad } f = \begin{pmatrix} -2x \\ -2y \end{pmatrix}$,

  $\Rightarrow$ minimum $= \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

  ➡ START $= \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\Delta = \frac{1}{20}$ (using only this Δ-value!)

  $$\mathbb{X}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbb{X}_1 = \mathbb{X}_0 + \Delta \cdot \vec{u}(\mathbb{X}_0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{20} \cdot \begin{pmatrix} -2 \\ -2 \end{pmatrix} = \begin{pmatrix} 9/10 \\ 9/10 \end{pmatrix} = \begin{pmatrix} 3^2/10^1 \\ " \end{pmatrix}$$

  $$\mathbb{X}_2 = \ldots = \begin{pmatrix} 81/100 \\ 81/100 \end{pmatrix} = \begin{pmatrix} 3^4/10^2 \\ " \end{pmatrix}$$

  $$\mathbb{X}_3 = \ldots = \begin{pmatrix} 729/1000 \\ 729/1000 \end{pmatrix} = \begin{pmatrix} 3^6/10^3 \\ " \end{pmatrix}$$

  $$\mathbb{X}_N = \begin{pmatrix} 3^{2N}/10^N \\ " \end{pmatrix} \qquad \left(\text{e.g., } \mathbb{X}_{10} \approx \begin{pmatrix} .3487 \\ .3487 \end{pmatrix} \right)$$

  $\Rightarrow$ ADAPTIVE stepsize important when 'close' to minimum (where |grad f| is VERY small). USE RUNGE-KUTTA METHODS WITH ADAPTIVE STEPSIZE.

- 'Getting close' to the exact minimum of a function with very small gradient magnitudes close to the minimum makes a robust and efficient method necessary!

- (Above example: $\mathbb{X}_N - \mathbb{X}_{N+1} = \begin{pmatrix} 3^{2N}/10^N - 3^{2(N+1)}/10^{N+1} \\ " \end{pmatrix}$

  $= \begin{pmatrix} 3^{2N}/10^{N+1} \\ " \end{pmatrix} \Rightarrow$ distance between points decreases with increasing N)

  $\Rightarrow$ SLOW!