# Implementation approaches of Fixed-point iteration method for Systems of Nonlinear equations

D. A. Vinnik
Taras Shevchenko National University of Kyiv

Kyiv 2019

# Basic Concepts

Let we need to solve a system:

$$\begin{cases} f_1(x1, \dots, x_n) = 0 \\ \dots \\ f_n(x1, \dots, x_n) = 0 \end{cases}$$

Let us write it as operator:

$$F(\bar{x}) = 0,$$

$$\bar{x} = (x_1, \dots, x_n)^T$$

We shall rewrite it in the form:

$$\bar{x} = S(\bar{x}),$$

where $S(x) = x + \tau B^{-1} F(x)$,

$\tau$ – numeric parameter, $B - n \times n$ matrix.

Consider an iterative process:

$$\bar{x}^{k+1} = S(\bar{x}^k) \ \sim \ \bar{x}^{k+1} = \bar{x}^k + \tau_{k+1} B_{k+1}^{-1} F(\bar{x}^k), \qquad \text{[1; Pt. 2, Ch. 5, p. 208]}$$

when $B_{k+1} = E, \tau_{k+1} = \tau = const$

For such a case:

- $q = \max\limits_{\bar{x} \epsilon B_r(\bar{x}_0)} \|S'(\bar{x})\| = \max\limits_{\bar{x} \epsilon B_r(\bar{x}_0)} \|E + \tau F'(\bar{x})\|$ [2; Ch. 4, §24, p. 184]
- Such estimates of accuracy is fair

  $\|\bar{x}^* - \bar{x}^k\| \leq \frac{q}{1-q} \|\bar{x}^k - \bar{x}^{k-1}\|$, [3; Ch. 1, §1.5.3, p. 44]

  $\|\bar{x}^* - \bar{x}^k\| \leq \frac{q^k}{1-q} \|\bar{x}^1 - \bar{x}^0\|$, [1; Pt. 2, Ch. 5, p. 209]

  $\|\bar{x}^* - \bar{x}^k\| \leq 2q^k r$ [2; Ch. 4, §24, p. 180]
- Sufficient convergence condition:

  $0 \leq q < 1, r \geq \frac{\|S(\bar{x}^0) - \bar{x}^0\|}{1-q}$ [2; Ch. 4, §24, p. 180]
- Priority estimate for accuracy $\varepsilon$:

  $k = \max\{0, \left[\frac{1}{\ln q} \ln \frac{\varepsilon}{2r}\right] + 1\}$ [2; Ch. 4, §24, p. 181]

# Computer implementation questions

a) Norm selection

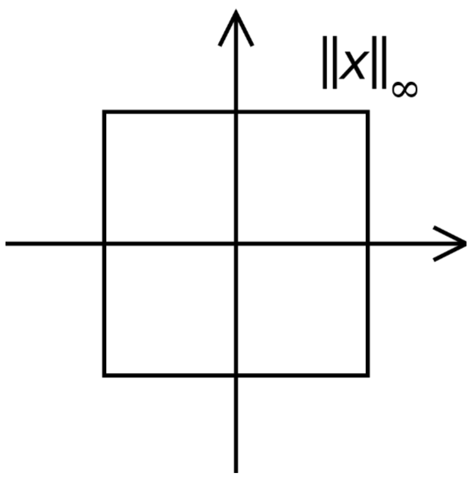
The most convenient norm is:

$\|A\|_\infty = \max_i \sum_j |a_{ij}|$

$\|\bar{x}\|_\infty = \max_i |x_i|$

Because $q$ is searched as max value and most of math packages uses box searching borders, which corresponds $\|\bar{x}\|_\infty = \max_i |x_i|$ norm.

$B_r(\bar{a}) = \{\bar{x}: \|\bar{x} - \bar{a}\|_\infty \leq r\}$




b) $\tau$ selection

Consider one equation case:

$x = s(x) = x + \tau f(x)$

$f'(x) < 0$ (1)

---

Define $\tau$ borders:

$q = \max_{x \epsilon B_r(\bar{x}_0)} |s'(x)| = \max_{x \epsilon B_r(\bar{x}_0)} |1 + \tau f'(x)|$

$q < 1$

$\max_{x \epsilon B_r(\bar{x}_0)} |s'(x)| < 1 \ \sim \ \max_{x \epsilon B_r(\bar{x}_0)} |1 + \tau f'(x)| < 1 \ \sim \ -1 < 1 + \tau \max_{x \epsilon B_r(\bar{x}_0)} f'(x) < 1$

$\sim \ -2 < \tau \max_{x \epsilon B_r(\bar{x}_0)} f'(x) < 0 \ \sim$

Let $M = -\max_{x \epsilon B_r(\bar{x}_0)} f'(x) = (1) = \max_{x \epsilon B_r(\bar{x}_0)} |f'(x)|$

$\sim \ 0 < \tau < \frac{2}{M}$

$\tau \in (0, \frac{2}{M})$ (2)

---

Let us search optimal $\tau$:

$x_{n+1} = x_n + \tau f(x_n)$

Rewrite it as:

$$\frac{x_{n+1} - x_n}{\tau} = f(x_n)$$

Introduce an error:

$e_n = x_n - x_*, e_{n+1} = x_{n+1} - x_*, x_n = x_* + e_n$

Rewrite:

$$\frac{e_{n+1} - e_n}{\tau} = f(x_* + e_n)$$

Mean value theorem:

$$f(x_* + e_n) = f(x_*) + f'(x_* + \theta z_n)e_n = 0 + f'(x_* + \theta z_n)e_n = f'(x_* + \theta z_n)e_n, \theta \epsilon (0,1) \qquad (3)$$

$$\frac{e_{n+1} - e_n}{\tau} = f'(x_* + \theta z_n)e_n \;\sim\; e_{n+1} - e_n = \tau f'(x_* + \theta e_n)e_n \;\sim$$

$$\sim\; e_{n+1} = e_n + \tau f'(x_* + \theta e_n)e_n \;\sim\; e_{n+1} = e_n\big(1 + \tau f'(x_* + \theta e_n)\big)$$

$$\sim\; |e_{n+1}| = \big|e_n\big(1 + \tau f'(x_* + \theta e_n)\big)\big| \;\sim\; |e_{n+1}| \le |e_n||1 + \tau f'(x_* + \theta e_n)|$$

$$\sim\; |e_{n+1}| \le |e_n| \max_{u \in [x_*,\ x_* + e_n]} |1 + \tau f'(u)|$$

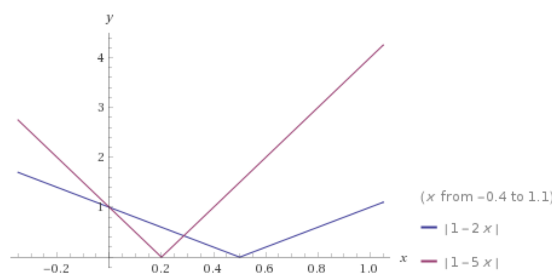
$0 < m < |f'(x)| < M$

$|e_{n+1}| \le |e_n| \max\{|1 - \tau M|, |1 - \tau m|\}$

$q(\tau) = \max\{|1 - \tau M|, |1 - \tau m|\}$

$$t_{opt} = \operatorname{argmin} q(t) = \frac{2}{M + m}$$

$$q_{opt} = q(t_{opt}) = \frac{M-m}{M+m}$$




$\{y = |1 - 2x|, y = |1 - 5x|\}$

---

Let us make analogue for multiple-equations case:

There is no such restriction on $\tau$ as (2) because we cannot disclose the norm as a module.

In addition, there is no natural mean value theorem for operators, but (3) is met with small in norm vectors $e_n$.

$|e_{n+1}| \le |e_n||1 + \tau f'(u)|$ will be $|\bar{e}_{n+1}| \le |\bar{e}_n| \max\limits_{\tau>0} \left\| E + \tau \max\limits_{\bar{x} \in [\bar{x}^*,\ \bar{x}^* + \bar{e}_n]} F'(x) \right\|$

Expand $\max\limits_{\tau>0} \left\| E + \tau \max\limits_{\bar{x} \in [\bar{x}^*,\ \bar{x}^* + \bar{e}_n]} F'(x) \right\|, diag(F') < 0$:

$$\max_{\tau>0} \max \begin{Bmatrix} \max\limits_{\bar{x} \in [\bar{x}^*,\ \bar{x}^* + \bar{e}_n]} \left( \left| 1 + \tau \dfrac{\partial f_1}{\partial x_1}(\bar{x}) \right| + \tau \sum\limits_{j \ne 1} \left| \dfrac{\partial f_1}{\partial x_j}(\bar{x}) \right| \right) \\ \max\limits_{\bar{x} \in [\bar{x}^*,\ \bar{x}^* + \bar{e}_n]} \left( \left| 1 + \tau \dfrac{\partial f_2}{\partial x_2}(\bar{x}) \right| + \tau \sum\limits_{j \ne 2} \left| \dfrac{\partial f_2}{\partial x_j}(\bar{x}) \right| \right) \\ \vdots \\ \max\limits_{\bar{x} \in [\bar{x}^*,\ \bar{x}^* + \bar{e}_n]} \left( \left| 1 + \tau \dfrac{\partial f_n}{\partial x_n}(\bar{x}) \right| + \tau \sum\limits_{j \ne n} \left| \dfrac{\partial f_n}{\partial x_j}(\bar{x}) \right| \right) \end{Bmatrix} =$$

$$
= \max_{\tau>0} \max \left\{ \begin{matrix} \max\limits_{\bar{x}} \left| 1 + \tau \frac{\partial f_1}{\partial x_1}(\bar{x}) \right| + \tau \sum\limits_{j \neq 1} \max\limits_{\bar{x}} \left| \frac{\partial f_1}{\partial x_j}(\bar{x}) \right| \\ \max\limits_{\bar{x}} \left| 1 + \tau \frac{\partial f_2}{\partial x_2}(\bar{x}) \right| + \tau \sum\limits_{j \neq 2} \max\limits_{\bar{x}} \left| \frac{\partial f_2}{\partial x_j}(\bar{x}) \right| \\ \vdots \\ \max\limits_{\bar{x}} \left| 1 + \tau \frac{\partial f_n}{\partial x_n}(\bar{x}) \right| + \tau \sum\limits_{j \neq n} \max\limits_{\bar{x}} \left| \frac{\partial f_n}{\partial x_j}(\bar{x}) \right| \end{matrix} \right\} =
$$

$$
= \max_{\tau>0} \max \left\{ \begin{matrix} \max\left\{ \left| 1 + \tau \max\limits_{\bar{x}} \frac{\partial f_1}{\partial x_1}(\bar{x}) \right|, \left| 1 + \tau \min\limits_{\bar{x}} \frac{\partial f_1}{\partial x_1}(\bar{x}) \right| \right\} + \tau \sum\limits_{j \neq 1} \max\limits_{\bar{x}} \left| \frac{\partial f_1}{\partial x_j}(\bar{x}) \right| \\ \max\left\{ \left| 1 + \tau \max\limits_{\bar{x}} \frac{\partial f_2}{\partial x_2}(\bar{x}) \right|, \left| 1 + \tau \min\limits_{\bar{x}} \frac{\partial f_2}{\partial x_2}(\bar{x}) \right| \right\} + \tau \sum\limits_{j \neq 2} \max\limits_{\bar{x}} \left| \frac{\partial f_2}{\partial x_j}(\bar{x}) \right| \\ \vdots \\ \max\left\{ \left| 1 + \tau \max\limits_{\bar{x}} \frac{\partial f_n}{\partial x_n}(\bar{x}) \right|, \left| 1 + \tau \min\limits_{\bar{x}} \frac{\partial f_n}{\partial x_n}(\bar{x}) \right| \right\} + \tau \sum\limits_{j \neq n} \max\limits_{\bar{x}} \left| \frac{\partial f_n}{\partial x_n}(\bar{x}) \right| \end{matrix} \right\} =
$$

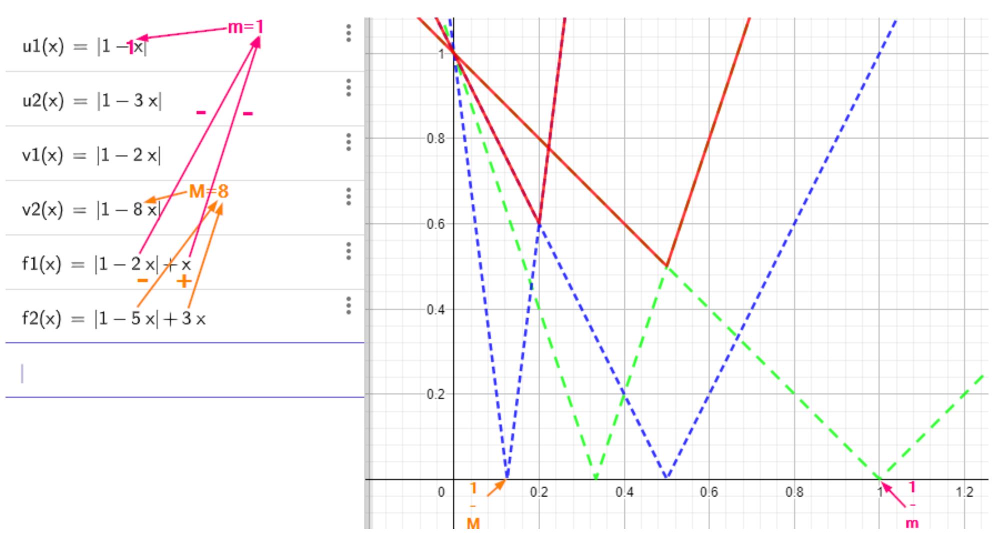
$$
= \max_{\tau>0} \max \left\{ \begin{matrix} \max\left\{ \left| 1 + \tau \left( \begin{matrix} \max \\ \min\limits_{\bar{x}} \end{matrix} \frac{\partial f_1}{\partial x_1}(\bar{x})) \begin{matrix} + \\ - \end{matrix} \sum\limits_{j \neq 1} \max\limits_{\bar{x}} \left| \frac{\partial f_1}{\partial x_j}(\bar{x}) \right| \right) \right| \right\} \\ \max\left\{ \left| 1 + \tau \left( \begin{matrix} \max \\ \min\limits_{\bar{x}} \end{matrix} \frac{\partial f_2}{\partial x_2}(\bar{x})) \begin{matrix} + \\ - \end{matrix} \sum\limits_{j \neq 2} \max\limits_{\bar{x}} \left| \frac{\partial f_2}{\partial x_j}(\bar{x}) \right| \right) \right| \right\} \\ \vdots \\ \max\left\{ \left| 1 + \tau \left( \begin{matrix} \max \\ \min\limits_{\bar{x}} \end{matrix} \frac{\partial f_n}{\partial x_n}(\bar{x})) \begin{matrix} + \\ - \end{matrix} \sum\limits_{j \neq n} \max\limits_{\bar{x}} \left| \frac{\partial f_n}{\partial x_j}(\bar{x}) \right| \right) \right| \right\} \end{matrix} \right\} = q(\tau)
$$

$$
M = \max_i \left( - \begin{matrix} \max \\ \min\limits_{\bar{x}} \end{matrix} \frac{\partial f_i}{\partial x_i}(\bar{x})) \begin{matrix} + \\ - \end{matrix} \sum_{j \neq 1} \max_{\bar{x}} \left| \frac{\partial f_i}{\partial x_j}(\bar{x}) \right| \right) > m
$$

$$
m = \min_i \left( - \begin{matrix} \max \\ \min\limits_{\bar{x}} \end{matrix} \frac{\partial f_i}{\partial x_i}(\bar{x})) \begin{matrix} + \\ - \end{matrix} \sum_{j \neq 1} \max_{\bar{x}} \left| \frac{\partial f_i}{\partial x_j}(\bar{x}) \right| \right) > 0
$$

$$
t_{opt} = \operatorname{argmin} q(t) = \frac{2}{M+m}
$$

$$
q_{opt} = q(t_{opt}) = \frac{M-m}{M+m}
$$



# References


1. *A. A. Samarsky, A. V. Gulin "Chislennyye metody" ("Numerical methods"). "Nauka", 1989.*
2. *E. A. Volkov "Chislennyye metody" ("Numerical methods"). "Lan'", 2008. ISBN 978-5-8114-0538-1.*
3. *Ya. S. Bugrov, S. M. Nikolsky "Vysshaya matematika. Differentsial'nyye uravneniya. Kratnyye integraly. Ryady. Funktsii kompleksnogo peremennogo" ("Higher Mathematics. Differential Equations, Multiple Integrals, Series, Theory of Functions of a Complex Variable"). "Drofa", 2004. ISBN 5-7107-8450-8.*