# $\ell_p$-Norm Minimization Method for Solving Nonlinear Systems of Equations

ANDRÉ A. KELLER
Laboratoire d'Informatique Fondamentale de Lille/ Section SMAC (UMR 8022 CNRS)
Université de Lille 1 Sciences et Technologies
Cité Scientifique, 59100 Lille
FRANCE
andre.keller@univ-lille1.fr

*Abstract:* Solving nonlinear systems of equations also refers to an optimization problem. Moreover, this equivalence can be interpreted as an optimal design problem. We must determine the design variables needed to reduce the deviation between an actual vector of valued functions and a vector of desired constants. This study focuses on these two characteristic features for solving nonlinear systems of equations. One variety of numerical two-dimensional systems with multiple solutions helps demonstrate the effectiveness of the transformation process.

*Key–Words:* Nonlinear system, least-squares-problem, regularized least-squares, $\ell_p$-norm, Newton's method

## 1 Introduction

Solving nonlinear systems of equations and optimization problems are major challenges in practical engineering (see [1, 2]). Indeed in an optimal design problem (see [3]-[7]), the quality of the design is measured by a norm of the deviation between an actual vector function of valued design variables $\mathbf{f}(\tilde{\mathbf{x}})$ and a vector of desired constants $\mathbf{b}$, i.e., $\mathbf{f}(\tilde{\mathbf{x}}) \approx \mathbf{b}$**.** The goal is to choose a vector of design variables $\mathbf{x}$ , such as the deviation is minimized.

Suppose we have to minimize the scalar multivariate function $f(\mathbf{x})$. A related problem is that of solving the system $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, where $\mathbf{g}$ is the gradient vector of $f$ with respect to $\mathbf{x}$, i.e., $\mathbf{g} \equiv \partial f/\partial \mathbf{x}$. Numerical techniques (e.g., Newton's method, Newton-Raphson, quasi Newton algorithms) usually solve these problems by iteration (see [3]-[6]). Let $\mathbf{G}(\mathbf{x})$ be the matrix of partial derivatives of $\mathbf{g}(\mathbf{x})$ with respect to $\mathbf{x}$. Using the root of the linear Taylor expansion about a current value $\mathbf{x}^{(k)}$ at iteration $k$, we get the new approximation $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{G}(\mathbf{x}^{(k)})^{-1}\mathbf{g}(\mathbf{x}^{(k)})$. To ensure that $\mathbf{G}$ is positive definite for a minimization problem, we solve $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{G}(\mathbf{x}^{(k)}) + \lambda_k \mathbf{I})^{-1}\mathbf{g}(\mathbf{x}^{(k)})$, where $\lambda_k$ is sufficiently large so that $f(x^{(k+1)}) < f(x^{(k)})$ (see also [7]).

The vector function $\mathbf{g}$ might not be the gradient of some objective function $f$. In that case, a relation can be stated between the solution of a system of nonlinear equations $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ where $\mathbf{g} : \mathbb{R}^n \mapsto \mathbb{R}^n$ , and the minimization of the scalar function $f(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T \mathbf{g}(\mathbf{x})$, for which the functional value at the minimum solution is zero. In that case, the Newton step is replaced by the line search strategy [1], i.e., $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \mathbf{G}(\mathbf{x}^{(k)})^{-1}\mathbf{g}(\mathbf{x}^{(k)})$, where $\alpha_k \in (0, 1)$.

More generally, consider the nonlinear system $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^m, (m > n)$. Since it is not possible to find an exact solution for that overdetermined system, one possibility is to seek a best least-squares solution, i.e., to find $\hat{\mathbf{x}}$ such that $\|\mathbf{f}(\mathbf{x})\|_2^2$ is minimized [8]. Approaches of least-squares trace back to the estimation methods from astronomical observation data. They were initially proposed by P.S. Laplace, L. Euler in the 1750s, by A.M. Legendre in 1805 and C.F. Gauss in 1809 to estimate motion orbits of planets (see [8]). Aaron [9] in 1956 uses least-squares in the design of physical systems. Gradient methods as steepest descents (suggested by Hadamar in 1907) has been used to solve systems of algebraic equations and least-squares problems (see [10]). In 1949, Booth [11] applied the steepest descents to the solution of nonlinear systems of equations.

The generalization of least-squares into $p$th least approximation $(p > 2)$ is presented by Tomes [12] with application to electric circuits and systems. In 1972, Bandler and Charalambous [13, 14] applied least $p$-th approximation to design problems and proved higher performances.

This application-oriented article is focused on the resolution of nonlinear systems of equations, using $p$th least approximations $(p \geq 1)$ [2]. Let a system of

[1]Recall that a suboptimization is implemented at each iteration step, minimizing $f(\mathbf{x}^{(k+1)})$ with respect to $\alpha_k$ [7].

[2]In the heat transfer process by [15], the resulting nonlinear system is solved using the Taylor linear approximation-based

nonlinear differentiable equations be

$$f_i(\mathbf{x}) = 0, \ i = 1, ..., m, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n, \ n < m$. Forming the semidefinite functional $J(\mathbf{x}) = \sum_{i=1}^{m} f_i^2(\mathbf{x})$, we seek the minimum of $J(\mathbf{x})$. If some minimum $\hat{\mathbf{x}}$ verifies $J(\hat{\mathbf{x}}) = 0$ then $\mathbf{x} = \hat{\mathbf{x}}$ is the solution of (1). If for every $\tilde{\mathbf{x}}$, we have $J(\tilde{\mathbf{x}}) > 0$, then $\mathbf{x} = \tilde{\mathbf{x}}$ are the least-squares solutions of (1). The interest of such optimization approaches is shown for different examples, for which multiple global minima exist. In such cases of equivalent global optimization problems, evolutionary algorithms may be use as in [16, 17].

This article is organized as follows. Section 1 presents historical context of this study and defines the objective . Section 2 introduces to the basic elements of least-squares and $p$th least approximations. Section 3 solves different examples for which the difficulties increase. Section 4 treats cases for which regularized systems are considered, due to the necessity of obtaining not too large parameters. The concluding remarks are in Section 5. Finally nonlinear systems and optimization problem are compared with more numerical examples in the appendix.

# 2 $\ell_p$-Norm Approximation Method

The distance approach refers to an approximation problem for which different norms may be chosen, such as the $\ell_p$ norms. A standard Newton's method is used to find a solution locally for smooth objective functions.

## 2.1 Basics, Notation and Definitions

The $\ell_p$-norm of the real vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{x}\|_p \triangleq (\sum_{i=1}^{n} |x_i|^p)^{1/p}$$

for $p \in \mathbb{N}_+$. The usual $\ell_p$-norms are: the $\ell_1$ (also called, the "Manhattan" or "city block" norm) for $p = 1$ , the Euclidean $\ell_2$ (also called, the "root energy" or "least-squares" norm) for $p = 2$ and the Chebyshev $\ell_\infty$ (also called, the "infinity", "uniform" or "supremum" norm) for $p \to \infty$. Figure 1 (a)-(b) picture unit balls centered at zero $\mathcal{B}_1(\|.\|)$ i.e.,$\mathbf{x} : \|\mathbf{x}\| = 1$, in the 3D and 2D spaces respectively.

Newton method, for which the implementation of numerically algorithms is presented.

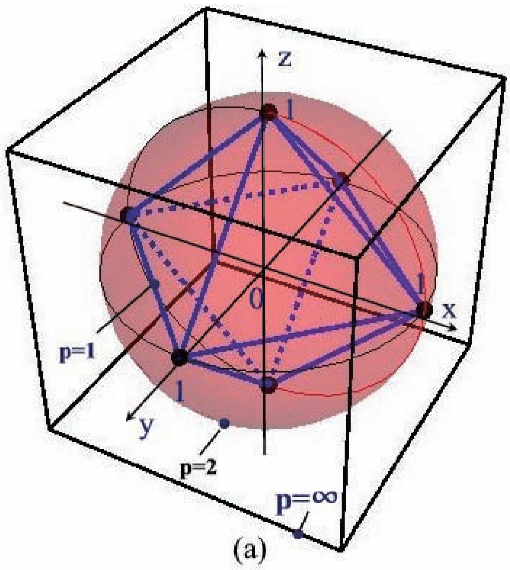



Figure 1: (a) 3D picture of unit balls in $\ell_p$ norm in $\mathbb{R}^3$ for $p \in \{1, 2, \infty\}$;

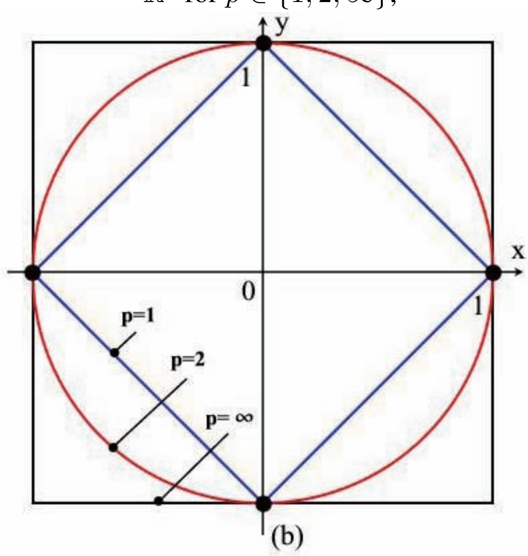



Figure 1: (b) 2D picture of unit balls in $\ell_p$ norm in $\mathbb{R}^2$ for $p \in \{1, 2, \infty\}$

An $\ell_p$-norm of a vector for an even $p$ is a differentiable function of its components. The function corresponding to an infinity norm $\ell_\infty$ is not differentiable. It can be shown that [18]

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_2 \geq \|\mathbf{x}\|_\infty.$$

Let $V$ be a vector space and $< .,. >$ an inner product on $V$. We define $\|\mathbf{v}\| = \sqrt{< \mathbf{v}, \mathbf{v} >}$. By the properties of an inner product, a norm satisfies

*(i)* $\|\mathbf{v}\| \geq 0$ with $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$,

*(ii)* $\|\alpha \ \mathbf{v}\| = |\alpha|.\|\mathbf{v}\|$ for $\alpha \in \mathbb{R}, \mathbf{v} \in V$ and

*(iii)* the triangle inequality or subadditivity $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$, for $\mathbf{v}, \mathbf{w} \in V$.

The basic norm approximation problem for an overdetermined system is [3, 19]

$$\text{minimize}_{\ \mathbf{x}} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2,$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \gg n$, $\mathbf{b} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$. The norm $\|.\|_2$ is on $\mathbb{R}^m$. A solution is an approximate solution of $\mathbf{Ax} \approx \mathbf{b}$ in the norm $\|.\|$ . The vector $\mathbf{r} = \mathbf{Ax} - \mathbf{b}$ is the residual for the problem. For Boyd and Vandenberghe [3] pp.291-292, four interpretations are possible for this problem, i.e., an interpretation in terms of regression, an interpretation in terms of parameter estimation, a geometry interpretation and an optimal design interpretation. Thus, in the regression problem, we can consider that the norm (i.e., a deviation measure) aims at approximating the vector $\mathbf{b}$ by a linear combination

$$\mathbf{Ax} = \sum_{j=1}^{n} \mathbf{a}^j x_j$$

of columns of $\mathbf{A}$. In the design interpretation, the $x_i$'s are the design variables to be determined. The goal is to choose a vector of design variables that approximates the target results $\mathbf{b}$ , i.e., $\mathbf{Ax} \approx \mathbf{b}$. The residual vector $\mathbf{r} = \mathbf{Ax} - \mathbf{b}$ expresses the deviation between actual and target results.

## 2.2 Least-Squares and Minimax Approximation Problems

We present two of the four possible interpretations, the regression problem and the optimal design problem. The norms used for this presentation are the Euclidean and the Chebyshev norms, for which we give the formulation in terms of an approximation problem [20].

• **Least-squares approximation problem**

By using the common Euclidean norm $\ell_2$, we obtain the following approximation problem

$$\text{minimize }_{\mathbf{x}} \quad f(\mathbf{x}) \equiv \|\mathbf{Ax} - \mathbf{b}\|_2^2,$$

where the objective $f(\mathbf{x})$ is the convex quadratic function:

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - 2\mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b}.$$

From the optimality conditions $\nabla f(\mathbf{x}) = 0$, we deduce that a solution point minimizes $f(\mathbf{x})$, if and only if, it satisfies the normal equations

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}.$$

Assuming independent columns for $\mathbf{A}$, a unique solution is achieved for

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}.$$

• **Minimax approximation problem**

By using the Chebyshev norm $\ell_\infty$, the approximation problem is

$$\text{minimize }_{\mathbf{x}} \quad \|\mathbf{Ax} - \mathbf{b}\|_\infty.$$

We may also write

$$\text{minimize }_{\mathbf{x}} \quad \max\{|\mathbf{A}_1\mathbf{x} - b_1|, \ldots, |\mathbf{A}_m\mathbf{x} - b_m|\},$$

where the $\mathbf{A}_i$'s are the rows of $\mathbf{A}$. Let $\alpha \in \mathbb{R}$, we have $|\mathbf{A}_i\mathbf{x} - b_i| \le \alpha$ for all $i = 1, \ldots, m$. Letting a vector of $m$ ones $\mathbf{e}_m = (1, 1, \ldots, 1)^T$, the condition is also written $|\mathbf{Ax}-\mathbf{b}| \le \alpha\, \mathbf{e}_m$. Therefore, an equivalent linear programming problem for the minimax problem is

$$\begin{cases} \text{minimize} & \alpha \\ \text{subject to} & |\mathbf{Ax} - \mathbf{b}| \le \alpha\, \mathbf{e}_m. \end{cases}$$

## 2.3 Nonlinear $\ell_p$-Norm Approximation Problem

We introduce to the the norm equivalence of quadratic objective functions and propose a formulation for generalized nonlinear systems.

• **Quadratic objective function norm equivalence**

Let the problem

$$\text{minimize }_{\mathbf{x}} \quad \mathbf{x}^T\mathbf{Qx}, \tag{2}$$

where $\mathbf{Q}$ is a $n \times n$ symmetric semidefinite matrix. Then, there is a matrix [3] $\mathbf{H}$, such as $\mathbf{Q} = \mathbf{HH}^T$, $\mathbf{H} \in \mathbb{R}^{n\times p}$, $p \ll n$. The reformulation of (2) is

$$\text{minimize }_{\mathbf{x}} \quad \|\mathbf{H}^T\mathbf{x}\|^2, \tag{3}$$

for which the storage and evaluation costs are more attractive with $n \times p$ for the reformulation (3) than $(1/2)n^2$ in the former formulation (2). In models of the covariance, we assume the generalization according to which $\mathbf{Q} = \mathbf{D} + \mathbf{HH}^T$, where $\mathbf{D}$ is a diagonal positive semi-definite matrix [21].

• **Generalized nonlinear approximation problem**

The nonlinear n-dimensional system $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$, has a solution when the scalar function $g(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x})$ has the minimum value of zero. Let the $f_i(\mathbf{x}), i = 1, \ldots, n$ be continuous component functions in the domain $\mathbf{x} \in X \subseteq \mathbb{R}^n$, we wish to determine the solution $\hat{\mathbf{x}} \in X$ for an initial approximation $\mathbf{x}^0$ . The first-order Taylor series approximation of the vector function $\mathbf{f}(\mathbf{x})$ is

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}^0) + \mathbf{J}(\mathbf{x}^0)\Delta\mathbf{x} + \mathcal{O}_2(\Delta\mathbf{x}),$$

where $\mathbf{J}$ is a $n \times n$ Jacobian matrix, $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^0$ denotes a correction error vector and $\mathcal{O}_2(\Delta\mathbf{x})$ is a negligible remainder for higher terms.

[3] The matrix $\mathbf{H}$ may be the Cholesky factor.

**Example 1** *Let the vector function*

$$f(x) = \begin{pmatrix} 6 + 20x_1 + 4x_1^2 + \frac{1}{5}x_2^2 \\ \\ 10 + 2x_1 - 4x_2 + \frac{1}{4}x_1x_2^2 \end{pmatrix}.$$

The linear approximation about one of the four solutions of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ (see Table 1) e.g., with $\hat{\mathbf{x}}^4 = (-0.3767, 2.1979)$, is

$$\tilde{\mathbf{f}}(\mathbf{x}) = \begin{pmatrix} 4.4663 + 16.9865x_1 + 0.8792x_2 \\ \\ 10.9099 + 3.2077x_1 - 4.4140x_2 \end{pmatrix}.$$

For Example 1, the scalar function $g(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x})$ is pictured in Figure 2 (b).

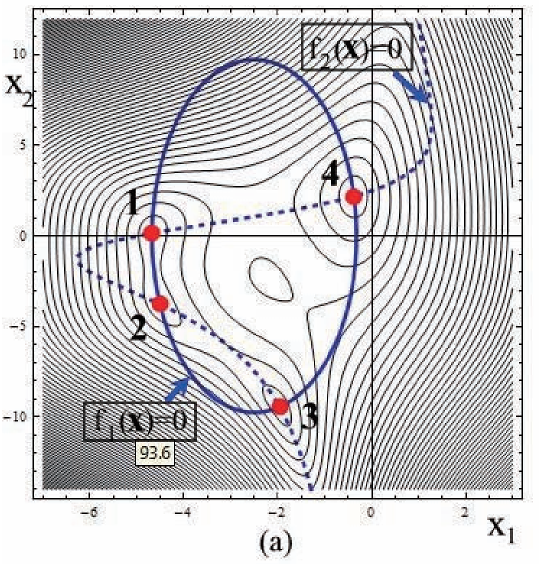



Figure 2: (a) Four minimum solutions of the system $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ in Example 1;

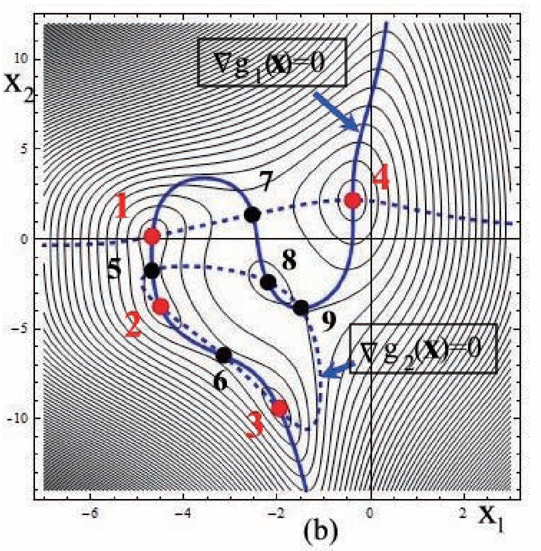



Figure 2: (b) Nine stationary points from the system $\nabla\mathbf{g}(\mathbf{x}) = 0$ in Example 1

The gradient $\nabla g(\mathbf{x})$ is given by

$$\begin{pmatrix} 280 + 904x_1 - 16x_2 + 480x_1^2 + 13x_2^2 \\ +64x_1^3 - 2x_2^3 + \frac{26}{5}x_1x_2^2 + \frac{1}{8}x_1x_2^4 \\ \\ -80 - 16x_1 + \frac{184}{5}x_2 + 26x_1x_2 + \frac{4}{25}x_2^3 \\ -6x_1x_2^2 + \frac{26}{5}x_1^2x_2 + \frac{1}{4}x_1^2x_2^3 \end{pmatrix}.$$

The four minimum solutions are presented in Table 1.

Table 1: Solutions for $f(\mathbf{x}) = 0$ in Example 1

| $\sharp$ | $\hat{x}_1$ | $\hat{x}_2$ | $\mathbf{f}(\hat{\mathbf{x}})$ |
|---|---|---|---|
| 1 | -4.6792 | 0.1535 | $4.5 \times 10^{-7}$ |
| 2 | -4.5090 | -3.7790 | $2.1 \times 10^{-7}$ |
| 3 | -1.9653 | -9.4489 | $1.5 \times 10^{-7}$ |
| 4 | -0.3767 | 2.1979 | $2.1 \times 10^{-10}$ |

Figure 2 (a) pictures the contours lines $g(\mathbf{x}) = C$ together with the four minimum solutions. The solutions are obtained by solving the system $\{f_1(\mathbf{x}) = 0, f_2(\mathbf{x}) = 0\}$. The stationary points solve the system of gradients

$$\Big\{\nabla_1 g(\mathbf{x}) = 0, \nabla_2 g(\mathbf{x}) = 0\Big\},$$

where $\nabla_i g(\mathbf{x}) = \partial g(\mathbf{x})/\partial x_i$, $i = 1, 2$ (see Figure 2 (b)) $\square$

The approximation of the quadratic function $g(\mathbf{x})$ has for expression

$$g(\mathbf{x}) = \mathbf{f}(\mathbf{x}^0)^T\mathbf{f}(\mathbf{x}^0) + (\Delta\mathbf{x})^T\mathbf{J}(\mathbf{x}^0)\mathbf{J}(\mathbf{x}^0)^T\Delta\mathbf{x} \\ +2\mathbf{f}(\mathbf{x}^0)^T\mathbf{J}(\mathbf{x}^0)\Delta\mathbf{x}.$$

A point $\mathbf{x}$ minimizes $g$, if and only if, $\mathbf{x}$ satisfies the normal equations

$$\mathbf{J}(\mathbf{x}^0)^T\mathbf{J}(\mathbf{x}^0)\Delta\mathbf{x} = -\mathbf{J}(\mathbf{x}^0)^T\mathbf{f}(\mathbf{x}^0).$$

Using the Newton's method, a step of the algorithm is (e.g., [4, 5])

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k\Big(\mathbf{J}(\mathbf{x}^{(k)})^T\mathbf{J}(\mathbf{x}^{(k)})\Big)^{-1}\mathbf{J}(\mathbf{x}^{(k)})\mathbf{f}(\mathbf{x}^{(k)}),$$

where $\alpha_k \in (0, 1)$ is a damping parameter. The steps of the Newton's approximation method in Table 2 are pictured in Figure 3. The exact values at iteration 20 are (a) $1.7\times10^{-7}$, (b) $2.1\times10^{-5}$, and (c) $-4.6\times10^{-5}$.

Table 2: Iteration steps of the Newton's method to Example 1

| $k$ | $x_1^{(k)}$ | $x_2^{(k)}$ | $\mathbf{f}(\mathbf{x}^{(k)})$ | $x_1^{(k)}-\hat{x}_1$ | $x_2^{(k)}-\hat{x}_2$ |
|---|---|---|---|---|---|
| 0 | 2 | -4 | 5695.0 | 2.377 | -6.198 |
| 1 | 1.373 | -3.309 | 2750.0 | 1.750 | -5.507 |
| 2 | 0.889 | -2.499 | 1331.6 | 1.266 | -4.696 |
| 3 | 0.527 | -1.546 | 636.7 | 0.904 | -3.744 |
| 4 | 0.256 | -0.491 | 286.6 | 0.633 | -2.688 |
| 5 | 0.045 | 0.529 | 112.1 | 0.422 | -1.669 |
| ⋮ | ⋮ | ⋮ | ⋮ | | |
| 10 | -0.356 | 2.150 | 0.173 | 0.0206 | -0.0480 |
| ⋮ | ⋮ | ⋮ | ⋮ | | |
| 20 | -0.377 | 2.198 | $0^{(a)}$ | $0^{(b)}$ | $-0^{(c)}$ |

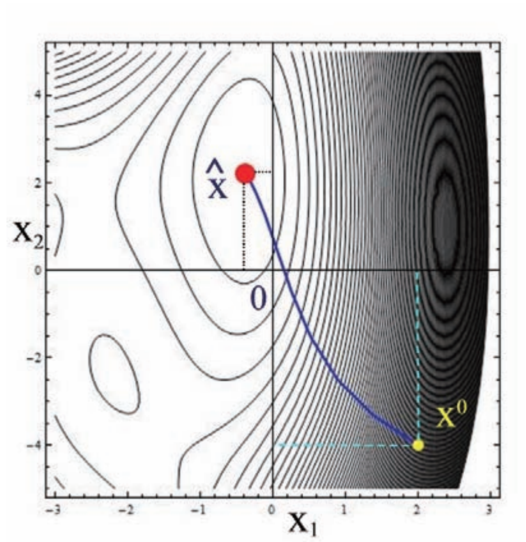



Figure 3: $\ell_2$-norm of $\mathbf{f}(\mathbf{x})$ in Example 1 by using the Newton's method

# 3 Numerical Examples

Numerical examples [4] consist in two-dimensional and higher dimensional systems. The two-dimensional examples are a polynomial system with few solutions and a trigonometric polynomial systems with much more real solutions. The common norms are compared. A three-dimensional example is reduced to a set of three two-dimensional subsystems by eliminating one of the variables.

## 3.1 Two-Dimensional Systems

**Example 2** *Let a residual vector be*

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} 25 - x_1^2 - x_2^2 \\ 5 + x_1 - x_2^2 \end{pmatrix}.$$

The two residual components equations $\mathbf{r}(\mathbf{x}) = \mathbf{0}$ yields three solution points at $\hat{\mathbf{x}}^1 = (-5, 0)^T$, $\hat{\mathbf{x}}^2 = (4, 3)^T$ and $\hat{\mathbf{x}}^3 = (4, -3)^T$. The $\ell_2$-norm approximation $\| \mathbf{r}(\mathbf{x}) \|_2$ is pictured in Figure 4 together with the three global minimum points for which the function values are zero. □

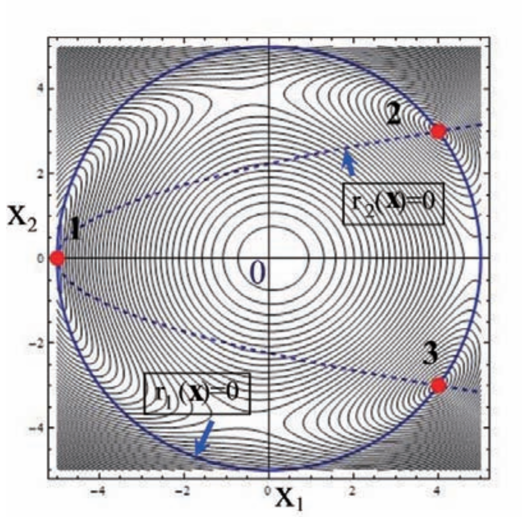



Figure 4: Polynomial Example 2 with $\ell_p$-norm approximation and the three solutions

**Example 3** *Let a residual vector be*

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} -3 + 4\cos(x_1) + 2\cos(x_2) \\ -1 + 2\sin(x_1) + \sin(x_2) \end{pmatrix}.$$

The two residual components in Figure 5 (a)-(b) show 18 solution points at the coordinates $\Big(a_1 \pm j2\pi, a_2 \pm j2\pi\Big)$ and $\Big(b_1 \pm j2\pi, b_2 \pm j2\pi\Big)$ for $j = 0, 1$, where $\mathbf{a} = (0.0658, 2.0894)^T$ and $\mathbf{b} = (1.1102, -0.9134)^T$. The $\ell_2$-norm approximation $\| \mathbf{r}(\mathbf{x}) \|$ is pictured in Figure 5 (b) together with the 18 global minimum points for which the function values are zero. □

[4] A collection of real-world nonlinear problems is presented in [22], e.g., an n-stage distillation column, the Bratu problem for nonlinear diffusion phenomena in combustion and semiconductors, the Chandrasekhar $H$-equation in radiative transfer problems, the elastohydrodynamic lubrication problem.

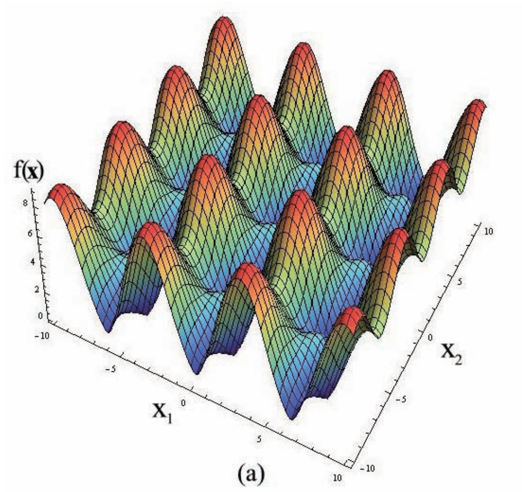



Figure 5: (a) 3D picture of the $\ell_2$-norm approximation for the polynomial trigonometric in Example 3;

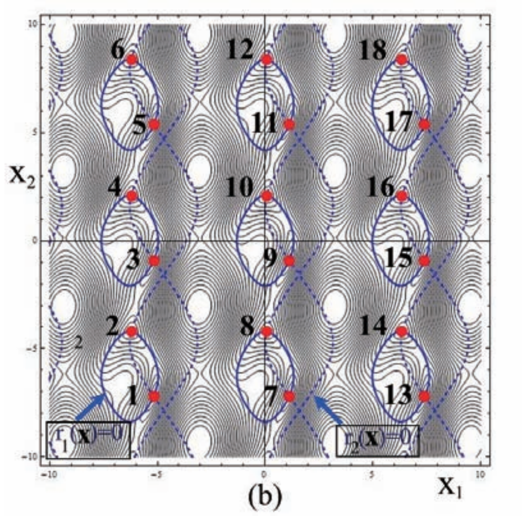



Figure 5: (b) Contours and the 18 minimum solution points

**Example 4** *Let a residual vector be*

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} x_1 - \sin(x_1 + 2x_2) - \cos(2x_1 - 3x_2) \\ x_2 - \sin(4x_1 - 3x_2) + \cos(x_1 + 2x_2) \end{pmatrix}.$$

The two residual components equations $\mathbf{r}(\mathbf{x}) = \mathbf{0}$ yield seven solution points at the coordinates given in Table 3.

The $\ell_2$-norm approximation $\parallel \mathbf{r}(\mathbf{x}) \parallel$ is pictured in Figure 6 (b) together with the seven global minimum points for which the function values are zero. □

Table 3: Solutions for $\mathbf{r}(\mathbf{x}) = \mathbf{0}$ in Example 4

| $\sharp$ | $\hat{x}_1$ | $\hat{x}_2$ | $\mathbf{f}(\hat{\mathbf{x}})$ |
|---|---|---|---|
| 1 | -1.8446 | -0.1325 | $2.1 \times 10^{-8}$ |
| 2 | -0.9409 | -0.0903 | $4.6 \times 10^{-8}$ |
| 3 | -0.0874 | 1.1303 | $5.3 \times 10^{-8}$ |
| 4 | 0.0523 | 1.7879 | $1.4 \times 10^{-8}$ |
| 5 | 0.4254 | -0.1431 | $8.6 \times 10^{-8}$ |
| 6 | 0.7231 | 0.9369 | $7.7 \times 10^{-8}$ |
| 7 | 1.3013 | 0.6190 | $2.1 \times 10^{-8}$ |

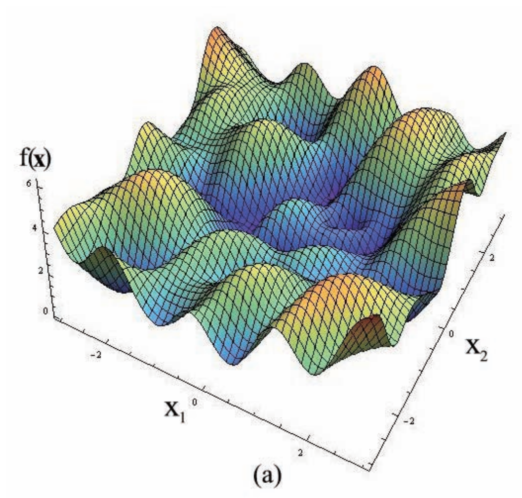



Figure 6: (a) 3D picture of the $\ell_2$ norm approximation for the trigonometric Example 4;

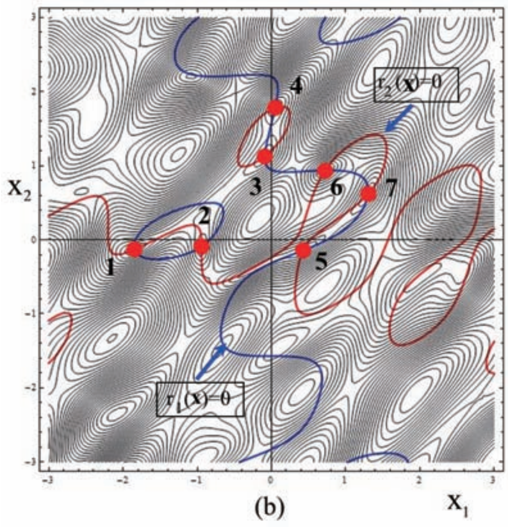



Figure 6: (b) Contours and the seven minimum solution points

## 3.2 Comparison of Usual Norms

**Example 5** *Let a residual vector be*

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} 1-4x_1+2x_1^2-2x_2^3 \\ -4+4x_2+x_1^4+4x_2^4 \end{pmatrix}.$$

An approximation problem for the three common norms $\ell_1$,$\ell_2$, and $\ell_\infty$ is pictured in Figure 7 (a) to (c) . The first Manhattan norm in Figure 7 (a) is defined as

$$\|\mathbf{r}(\mathbf{x})\|_1 = |r_1(\mathbf{x})| + |r_2(\mathbf{x})|.$$

The second Euclidean norm in Figure 7 (b) is defined as

$$\|\mathbf{r}(\mathbf{x})\|_2 = \left(|r_1(\mathbf{x})|^2 + |r_2(\mathbf{x})|^2\right)^{1/2}.$$

and is pictured with the two contour lines $r_1(\mathbf{x}) = 0$ and $r_2(\mathbf{x}) = 0$. The third Chebyshev norm in Figure 7 (c) is defined as

$$\|\mathbf{r}(\mathbf{x})\|_\infty = \max\{|r_1(\mathbf{x})|, |r_2(\mathbf{x})|\}$$

and is pictured with the exclusion lines along which we find the discontinuities of the function.

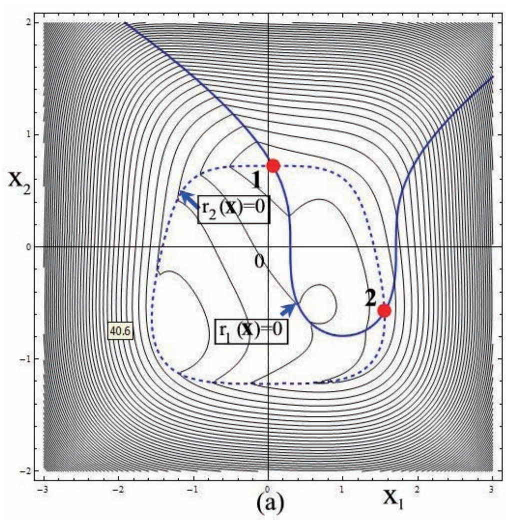



Figure 7: Example 5 with usual norm approximations: (a) Manhattan norm;

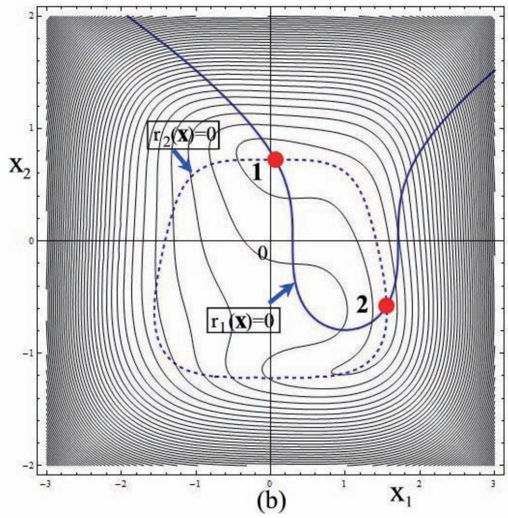



Figure 7: (b) Euclidean norm;

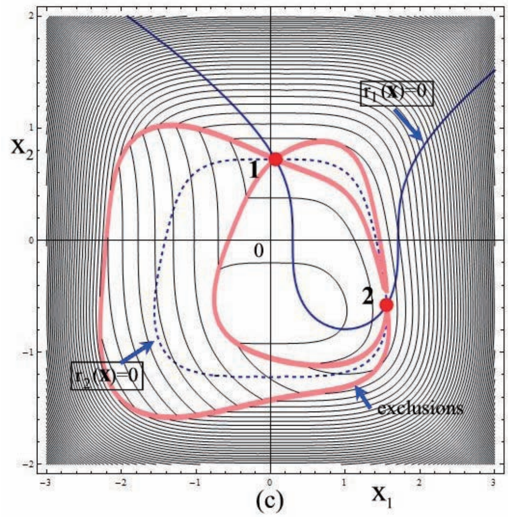



Figure 7: (c) Chebychev norm

The resulting functions of using the $\ell_1$ and $\ell_\infty$ norms are nonsmooth. Then, nonderivative optimization methods must be used (see [23] on optimization and nonsmooth analysis and [24] on the finite difference approximation of sparse Jacobian matrices in Newton methods).

## 3.3 Higher Dimensional Systems

A two-dimensional subsystem can be obtained from a higher dimensional system, by eliminating one or more of its variables. This procedure is illustrated by the following three- dimensional system for which three two-dimensional subsystems are deduced.

**Example 6** *Let a residual vector be*

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} -2+x_1-x_2+x_3 \\ x_1x_2x_3 \\ -1+2x_2+x_3 \end{pmatrix}.$$

• By eliminating $x_3$, we obtain the two-dimensional system

$$\tilde{\mathbf{r}}(x_1, x_2) = \begin{pmatrix} -1 + x_1 - 3x_2 \\ -x_2 - x_2^2 + 6x_2^3 \end{pmatrix}.$$

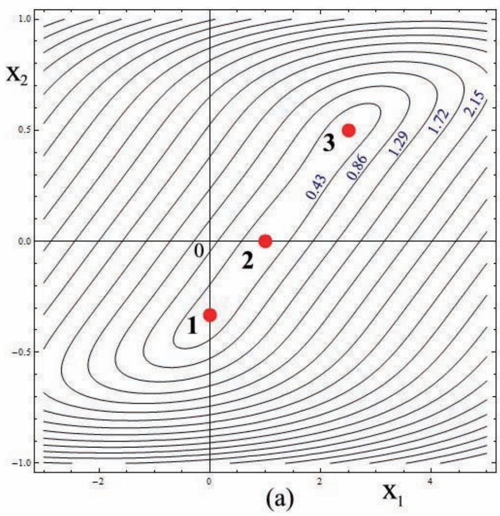
• By eliminating $x_2$, we obtain the two-dimensional system

$$\tilde{\mathbf{r}}(x_1, x_3) = \begin{pmatrix} -5 + 2x_1 + 3x_3 \\ 5x_3 - 8x_3^2 + 6x_3^3 \end{pmatrix}.$$

• By eliminating $x_1$, we obtain the two-dimensional system

$$\tilde{\mathbf{r}}(x_2, x_3) = \begin{pmatrix} -1 + 2x_2 + x_3 \\ 5x_3 - 8x_3^2 + 3x_3^3 \end{pmatrix}.$$

The three solutions for these subsystems are $\mathbf{x}^1 = \left(0, -\frac{1}{3}, \frac{5}{3}\right)^T$, $\mathbf{x}^2 = \left(1, 0, 1\right)^T$ and $\mathbf{x}^3 = \left(\frac{5}{2}, \frac{1}{2}, 0\right)^T$. The contour maps corresponding to these subsystems are pictured in Figure 8 (a), (b) and (c) respectively. □



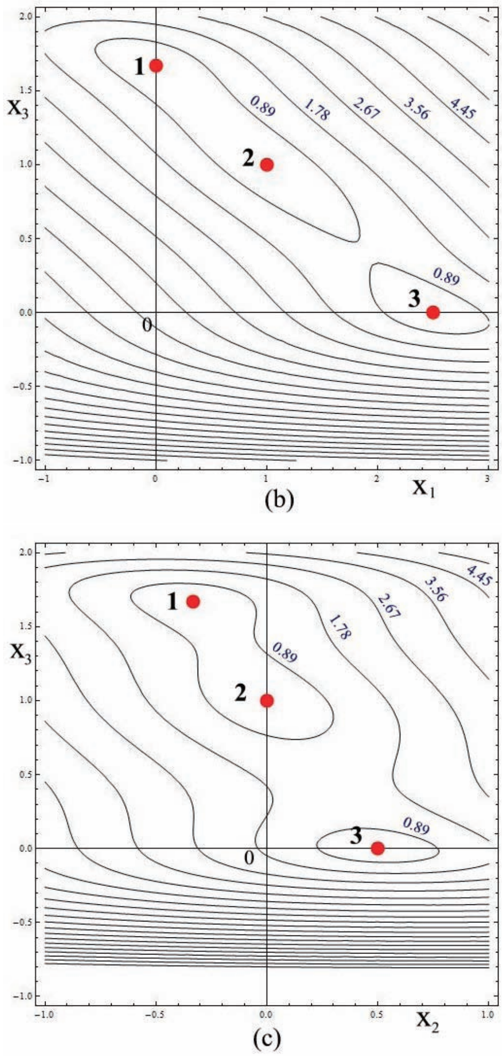




Figure 8: Contour maps of three two-dimensional subsystems in Example 6

# 4 Regularized Least-Squares

The norm approximation problem can be formalized for a regularized linear system of equations, for which one another objective is given.

In a regularized least-squares problem, the objectives are twofold. One first objective is to find the design variables $\mathbf{x} \in \mathbb{R}^n$ that gives a better fit. One another objective is to obtain not too large design variables. The vector optimization problem with respect to the cone $\mathbb{R}_+^2$ is

$$\text{minimize} \quad \mathbf{f}(\mathbf{x}) = \Big(f_1(\mathbf{x}), f_2(\mathbf{x})\Big)^T, \qquad (4)$$

where the functions $f_1$ and $f_2$ may represent two identical Euclidean norms (or different norms) measuring the fitting error and the size of the design vector, respectively (see[3, pp. 184-185]).

### 4.1 Tikhonov Regularization

Let the deterministic [5] overdetermined system be $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \gg n$, and $\mathbf{b} \in \mathbb{R}^m$ (e.g., in some data fitting problems). We retain quadratic measures for the size of the residuals $\mathbf{Ax} - \mathbf{b}$ and for that of $\mathbf{x}$. The problem (4) is to minimize the two squared norms, so that

$$\text{minimize} \quad \mathbf{f}(\mathbf{x}) \equiv \left(\|\mathbf{Ax}-\mathbf{b}\|_2^2, \|\mathbf{x}\|_2^2\right).$$

For both Euclidean norms, the unique Pareto optimal point is given by $\tilde{\mathbf{x}} = \mathbf{A}^\dagger\mathbf{b}$, where $\mathbf{A}^\dagger$ denotes the pseudoinverse of $\mathbf{A}$, i.e., $\mathbf{A}^\dagger = \lim_{\varepsilon\to 0}(\mathbf{A}^T\mathbf{A} + \varepsilon I)^{-1}\mathbf{A}^T$ at $\varepsilon > 0$. Expanding $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, and scalarizing with strictly positive $\lambda_i$'s for $i = 1, 2$, the minimization problem is expressed by

$$\text{minimize}_{\ \mathbf{x}\in\mathbb{R}^n_+} \quad \left\{\mathbf{x}^T\left(\lambda_1\mathbf{A}^T\mathbf{A} + \lambda_2\mathbf{I}\right)\mathbf{x} - 2\lambda_1\mathbf{b}^T\mathbf{Ax} + \lambda_1\mathbf{b}^T\mathbf{b}\right\}.$$

The minimum solution point $\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}$ where $\mu = \lambda_2/\lambda_1$ is Pareto optimal for any $\mu > 0$ (see also this Tikhonov regularization technique by [3, 25]).

### 4.2 $\ell_1$-norm regularization

Different norms are used in preference for the two objectives in practical applications (e.g., image restoration) as in [26, 27]. The common method is

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{Ax}-\mathbf{b}\|_2 + \gamma\|\mathbf{x}\|_1,$$

where $\gamma \in (0, \infty)$. In this formulation, the cost of using large values is a penalty added to the cost of missing the goal specification. The regularization technique overcomes an ill-conditioned matrix $\mathbf{A}$ of data [26]. An optimal trade-off curve for a regularized least-squares problem may be determine, as in [3, p. 185].

## 5 Conclusion

Solving nonlinear systems using $\ell_p$-norms appears to be an effective method of resolution. The initial problem is transformed into an optimization problem for which we are looking for the zero global minimum solutions. However, the city-block $\ell_1$-norm and the uniform $\ell_\infty$-norm produce nonsmooth objective functions, for which nonderivative optimization techniques are helpful. This optimization method is applied to a variety of small size systems of nonlinear equations. This approach is well suited, specially, for trigonometric polynomial systems with multiple solutions.

[5] In presence of noisy observations, the system becomes $\mathbf{b} = \mathbf{Ax} + \xi$, where $\xi$ denotes a white Gaussian noise vector.

## 6 Appendix: Nonlinear Systems and Optimization Problem

### A.1 Problem Equivalence

Solving a nonlinear system of equation and optimizing a multivariate function include multiple similarities as in Table 4. The initial formulation of the two problems is different. In the first problem, the nonlinear system $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ consists in a vector function $\mathbf{g} : \mathbb{R}^n \mapsto \mathbb{R}^n$, while a multivariate scalar function is optimized in the second problem. In both cases, a Taylor series expansion is used. However, it is a linear approximation for the nonlinear system, but a quadratic approximation for the optimization. Similar assumptions are made with regards to aspects such as smoothness functions, negligible remainders, inversion of the squared Jacobian and Hessian matrix, respectively. Differences are in the formula for the Newton iteration step.

### A.2 Numerical Examples

Two examples illustrate the two procedures. In Example 7 the nonlinear system of equation $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ allows to determine the stationary points of the multivariate function $f(\mathbf{x})$ in the minimization problem. The connection is different in Example 8 since the objective function of the minimization problem is expressed as $f(\mathbf{x}) = \mathbf{r}(\mathbf{x})^T\mathbf{r}(\mathbf{x})$ , for which the residual error function satisfies $\mathbf{r}(\mathbf{x}) = \mathbf{0}$. The two-dimensional Rosenbrock's test function is used for this example.

**Example 7** *Let the bivariate function*

$$f(\boldsymbol{x}) = 3 - 3x_1 - 2x_2 + x_2^2 + x_1^3,$$

*where* $\boldsymbol{x} \in [-3, 3] \times [-2, 4]$.

The first iterations of the nonlinear system problem and that of the optimization problem are in Table 5. The gradient is $\mathbf{g}(\mathbf{x}) = \left(-3 + 3x_1^2, -2 + 2x_2^2\right)^T$. The solutions of the nonlinear system $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ are the stationary points $(-1, 1)^T$ and $(1, 1)^T$ for which the function values are respectively $4$ and $0$.

Table 4: Nonlinear system problem solving and optimization problem: a comparison

| Nonlinear System Solving Problem | Optimization Problem |
|---|---|
| **1. Formulation** | |
| • Find $\mathbf{x}$ such that $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{g} : \mathbb{R}^n \mapsto \mathbb{R}^n$. | • minimize $_{\mathbf{x}}$ $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^n$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$. |
| **2. Approximation** | |
| • Linear Taylor series expansion about $\mathbf{x}^0$, i.e., $\mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{x}^0) + \mathbf{J}(\mathbf{x}^0)^T \Delta\mathbf{x} + \mathcal{O}_2(\Delta\mathbf{x})$, where $\mathbf{J}(\mathbf{x}^0) \equiv \nabla\mathbf{g}(\mathbf{x}^0)$ is the Jacobian matrix at $\mathbf{x}^0$ and $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^0$, the correction vector. | • Quadratic Taylor series expansion about $\mathbf{x}^0$, i.e. $f(\mathbf{x}) = f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0)^T \Delta\mathbf{x} + \frac{1}{2}(\Delta\mathbf{x})^T \mathbf{H}(\mathbf{x}^0)\Delta\mathbf{x} + \mathcal{O}_3(\Delta\mathbf{x})$, where $\mathbf{H}(\mathbf{x}^0) \equiv \nabla^2 f(\mathbf{x}^0)$ is the Hessian matrix at $\mathbf{x}^0$ and $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^0$, the correction vector. |
| **3. Assumptions** | |
| • Smoothness of the functions $g_i(\mathbf{x}), i = 1, \ldots, n$ <br> • Negligible remainder $\mathcal{O}_2(\Delta\mathbf{x})$ for higher terms <br> • Inverse of the squared Jacobian $\mathbf{J}(\mathbf{x}^0)^T \mathbf{J}(\mathbf{x}^0)$. | • Smoothness of the functions $f(\mathbf{x})$ <br> • Negligible remainder $\mathcal{O}_3(\Delta\mathbf{x})$ for higher terms <br> • Inverse Hessian and positive definite Hessian for a minimum. |
| **4. Normal equation** | |
| $\mathbf{J}^T \mathbf{J} \Delta\mathbf{x} = -\mathbf{J}\mathbf{g}(\mathbf{x})$. | $\mathbf{H}\Delta\mathbf{x} = -\nabla f(\mathbf{x})$. |
| **5. Newton's iteration step** $k = 0, 1, \ldots$ | |
| $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \left(\mathbf{J}(\mathbf{x}^{(k)})^T \mathbf{J}(\mathbf{x}^{(k)})\right)^{-1} \mathbf{J}(\mathbf{x}^{(k)}\mathbf{g}(\mathbf{x}^{(k)})$. | $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \mathbf{H}(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$, where $\alpha_k \in (0, 1)$ is the step size. |

The global minimum is $(1, 1)^T$ at which the Hessian $\mathbf{H} = \begin{pmatrix} 6x_1 & 0 \\ 0 & 2 \end{pmatrix}$ is positive definite, and for which the function value is zero.

**Example 8** *The bivariate Rosenbrock's function is*

$$f(\boldsymbol{x}) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2,$$

*where* $\boldsymbol{x} \in [-1.5, 1.5]^2$.

The residual vector function for the Rosenbrock's function is

$$\mathbf{r}(\mathbf{x}) = \begin{pmatrix} 1 - x_1 \\ \\ 10(x_2 - x_1^2 \end{pmatrix},$$

such as $f(\mathbf{x}) = \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x})$. The iterations of the Newton's method are in Table 6.

Table 6: Newton's method in Example 8

| $k$ | $x_1^{(k)}$ | $x_1^{(k)}$ | $f(\mathbf{x}^{(k)})$ |
|---|---|---|---|
| 0 | -1 | 1 | 4 |
| 1 | 0 | -1 | 101 |
| 2 | 0.00248 | -0.5 | 25.9956 |
| 3 | 0.00742 | -0.25 | 7.2365 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 38 | 0.9999 | 0.9999 | $6 \times 10^{-13}$ |
| 39 | 1. | 0.9999 | $1.5 \times 10^{-13}$ |
| 40 | 1 | 1 | $63.8 \times 10^{-14}$ |

*References:*

[1] S. S. Rao, *Engineering Optimization: Theory and Practice* (third edition), J. Wiley & Sons, Hoboken, N. J. 2009

Table 5: Nonlinear system and optimization problem in Example 7

| k | Nonlinear system of gradients | | | Optimization problem | | |
|---|---|---|---|---|---|---|
| | $x_1^{(k)}$ | $x_2^{(k)}$ | $f(\mathbf{x}^{(k)})$ | $x_1^{(k)}$ | $x_2^{(k)}$ | $f(\mathbf{x}^{(k)})$ |
| 0 | 2 | 3 | 8 | 2 | 3 | 8 |
| 1 | 1.25 | 1 | 0.2031 | 1.625 | 2 | 2.4160 |
| 2 | 1.025 | 1 | 0.00189 | 1.3726 | 1.5 | 0.7182 |
| 3 | 1.0003 | 1. | $2.79\times10^{-7}$ | 1.2116 | 1.25 | 0.2063 |
| 4 | 1. | 1. | $6.44\times10^{-15}$ | 1.1150 | 1.125 | 0.0568 |
| 5 | 1. | 1. | $-2.22\times10^{-16}$ | 1.0605 | 1.0625 | 0.0151 |
| 6 | 1. | 1. | 0 | 1.0311 | 1.0313 | 0.0039 |
| 7 | 1. | 1. | 0 | 1.0158 | 1.0156 | $9.95\times10^{-4}$ |
| 8 | 1. | 1. | 0 | 1.0079 | 1.0078 | $2.51\times10^{-4}$ |
| 9 | 1. | 1. | 0 | 1.0039 | 1.0039 | $6.31\times10^{-5}$ |
| 10 | 1. | 1. | 0 | 1.002 | 1.002 | $1.6\times10^{-5}$ |

[2] R. Ravindran, K.M. Raqsdell and G.V. Reklaitis, *Engineering Optimization: Method and Applications* (second edition), J. Wiley & Sons, Hoboken, N. J. 2006

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge UK 2004

[4] P. A. Jensen and J.F. Bard, *Operations Research: Models, and Methods,* Wiley & Sons, Hoboken, N. J. 2003

[5] J. Nocedal and S.J. Wright, *Numerical Optimization* (second edition), Springer Science Business Media, New York 2006

[6] J. A. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical New Gradient-Based Algorithms*, Springer Science+Business Media, New York 2005

[7] G. K. Smyth, Optimization and nonlinear equations. In: P. Armitage and T. Colton (eds.), *Encyclopedia of Biostatistics*, pp. 3174–3180, J. Wiley & Sons, Chichester 1998

[8] C. Xu, Nonlinear least squares problems. In: C.A. Floudas and P.M. Pardalos (eds.), *Encyclopedia of Optimization* (second edition), pp. 2626–2630, Springer Science Business Media, New York, 2009

[9] M. R. Aaron, The use of least squares in system design, *IRE Trans. Circuit Theory* Vol.3, No. 4, 1956, pp. 224–231.

[10] J. W. Fischbach, Some applications of gradient methods. In: R.V. Churchill, E. Reissner, and A.H. Taub (eds.), *Numerical Analysis*, pp. 59–72, American Mathematical Society, Providence R.I., 1956

[11] A. D. Booth, An application of the method of steepest descents to the solution of systems of non-linear simultaneous equations, *Quart. J. Mech. Appl. Math.* Vol.2, No. 4, 1949, pp. 460–468.

[12] G. C. Tomes, Least pth approximation, *IEEE Trans. Circuit Theory*, Vol.16, No. 2, 1969, pp. 235–237.

[13] J. W. Bandler and C. Charalambous, Theory of generalized least *p*th approximation, *IEEE Trans. Circuit Theory*, Vol.19, No. 3, 1972, pp. 287–289.

[14] J. W. Bandler and C. Charalambous, Practical least *p*th optimization of networks, *IEEE Trans. Microwave Theory Tech.*, Vol.20, No. 12, 1972, pp. 834–840.

[15] C. Patrascioiu and C. Marinoiu, The applications of the nonlinear equations systems algorithms for the heat transfer processes, *Proc. 12th WSEAS Int. Conf. MAMECTIS'10*, Kantaoui, Sousse, Tunisia, May 3-6, 2010, pp.30-35.

[16] A. F. Kuri-Morales, Solution of simultaneous non-linear equations using genetic algorithms, *WSEAS Trans. Systems*, Vol.2, No. 1, 2003, pp. 44–51.

[17] N. E. Mastorakis, Solving nonlinear equtions via genetic algorithms, *WSEAS Trans. Inform. Sci. Appl.*, Vol.2, No. 5, 2005, pp. 455–459.

[18] A. Antoniou and W. S. Lu, *Practical Optimization. Algorithms and Engineering Applications* , Springer Science Business Media, New York 2007

[19] R. Fletcher, *Practical Methods of Optimization* (second edition), John Wiley & Sons, Chichester - New York -Brisbane 1987

[20] W. Forst and D. Hoffmann, *Optimization- Theory and Practice,* Springer Science Business Media, New York 2010

[21] E. D. Andersen, On formulating quadratic functions in optimization models, *MOSEK Technical Report TR-1-2013*, 2013.

[22] J. J. Moré, A collection of nonlinear model problems, In: E. L. Allgower and K. Georg (eds.), *Computational Solution of Nonlinear Systems of Equations*, pp. 723–762, American Mathematical Society, Providence, R.I. 1990

[23] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, P. A.1990

[24] T. J. Ypma, Finite difference approximation of sparse jacobian matrices in Newton-like methods. In: E.L. Allgower and K. Georg (eds.), *Computational Solution of Nonlinear Systems of Equations*, pp. 707–722, American Mathematical Society, Providence, R.I. 1990

[25] X.-S. Yang, *Engineering Optimization: An Introduction with Metaheuristic Applications*, John Wiley & Sons, Hoboken N. J. 2010, pp. 97–99.

[26] L. Wu and Z. Sun, New nonsmooth equations-based algorithms for $\ell_1$-norm minimization and applications, *J. Appl. Math.*, 2012, doi:10.1155/2012/139609

[27] Y. Xiao, Q. Wang and Q. Hu, Non-smooth equations based method for $\ell_1$-norm problems with applications to compressed sensing, *Nonlinear Anal.*, Vol.74, 2011, pp. 3570–3577.