Unconstrained Nonlinear Optimization Algorithms

Why study it?

Algorithms for the solution of constrained problems are based on algorithms for the solution of unconstrained problems (penalization methods, determination of search direction, etc.);

Basic Principle: Problem with n variables ➡ Problem with a single variable

$$\mathbf{x} = \mathbf{x}_0 + \alpha \mathbf{s}_0 \;\Rightarrow\; f(\mathbf{x}) = f(\mathbf{x}_0 + \alpha \mathbf{s}_0) = f(\alpha) \;\Rightarrow\; \text{Find } \alpha$$

Typical Procedure

- Find $\mathbf{s}_0$ at $\mathbf{x}_0$ that reduces the objective function
- Find α in the direction of $\mathbf{s}$ $\Rightarrow \mathbf{x}_0 + \alpha\mathbf{s}_0$ (unidimensional search)
- Verify convergence: if satisfied $\Rightarrow \mathbf{x} = \mathbf{x}^*$, stop, otherwise $\Rightarrow \mathbf{x}_{i+1} = \mathbf{x}$ and $i = i + 1$

Unconstrained Nonlinear Optimization Algorithms

Classification

- Zeroth-Order ➡ uses only objective function values. Used when the function is not differentiable or when the function is highly nonlinear, and therefore the derivatives are difficult to obtain with precision.
  Ex.: Powell's Conjugate Direction Method
- First-Order ➡ employs only the function value and its gradient
  Ex.: "Steepest Descent" and Conjugate Gradient Methods
- Second Order ➡ Employs the objective, the gradient and the Hessian of the function.
  Ex.: Newton's and Quasi-Newton Methods

Zeroth-Order Methods

3

Used when the function value is obtained with poor precision ➡ Determination of derivative (or gradient) values are not reliable.

Powell's Method of Conjugate Directions

Based on the minimization of a quadratic function considering linearly independent *Q*-conjugate directions. Quadratic function:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x} + \mathbf{c}$$

Q-conjugate directions: $\mathbf{s}_i^T\mathbf{Q}\mathbf{s}_j = 0$ for $i \neq j$

"If f is minimized along a direction **s**, then the minimum of f will occur at (or before) the n-th step regardless of the start point, provided rounding errors do not accumulate."

Powell ➡ convenient method for generating *Q*-conjugated, linearly independent directions. If the generated directions are not linearly dependents ➡ convergence to a minimum will not occur.

Powell's Method of Conjugate Directions

4

Procedure:

1. Minimize f along coordinate directions (univariate search), starting at $\mathbf{x}_0^k$ and generating points $\mathbf{x}_1^k, \ldots, \mathbf{x}_n^k$, where k is the cycle index;

2. After the univariate search is finished, find the index corresponding to the direction in which f presents the greatest reduction, from $\mathbf{x}_{m-1}^k$ to $\mathbf{x}_m^k$;

3. Calculate the "usual" direction $\mathbf{s}_p^k = \mathbf{x}_n^k - \mathbf{x}_0^k$ and find the α-value that minimizes f such that: $\mathbf{x} = \mathbf{x}_0^k + \alpha\mathbf{s}_p^k$

4. If

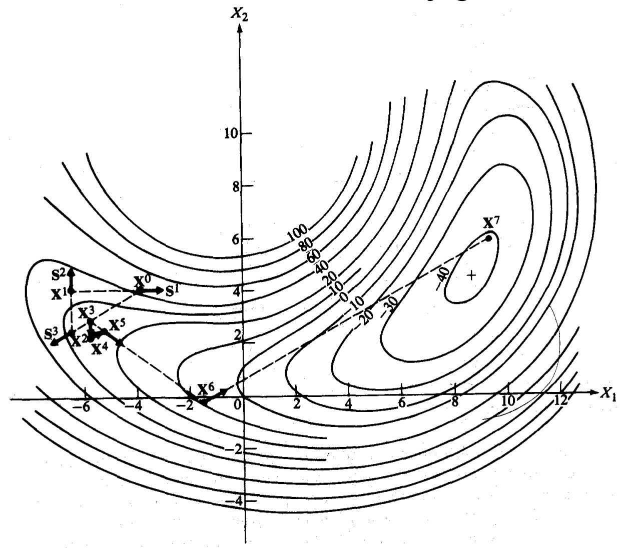
$$|\alpha| < \left[\frac{f(\mathbf{x}_0^k) - f(\mathbf{x}_0^{k+1})}{f(\mathbf{x}_{m-1}^k) - f(\mathbf{x}_m^k)}\right]^{\frac{1}{2}}$$

then use the same directions for the next univariate search. If the equation is NOT satisfied, then replace the m-th direction by the usual direction $\mathbf{s}_p^k$

5. Start a new univariate search with the directions obtained in step 4, and repeat steps 2, 3, e 4 until convergence, i.e., $\left\|\mathbf{x}_{k+1} - \mathbf{x}^k\right\| \leq \varepsilon$

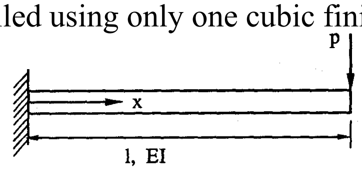
Powell's Method of Conjugate Directions

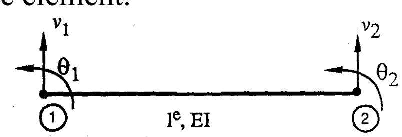
Interpretation of the Powell's Method of Conjugate Directions:



Example

Study problem: Determination of maximal deflection and rotation of the extremity of a beam after the minimization of its total potential energy, modelled using only one cubic finite element.





Finite element formulation:

$$v(\xi) = \left[\left(1-3\xi^2+2\xi^3\right) \quad l\left(\xi-2\xi^2+\xi^3\right) \quad \left(3\xi^2-2\xi^3\right) \quad l\left(-\xi^2+\xi^3\right)\right]\begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \end{Bmatrix}$$

where: $\xi = x/l$. The corresponding potential energy of the beam model is given by: $\Pi = \dfrac{EI}{2l^3}\displaystyle\int_0^l \left(\dfrac{d^2v}{d\xi^2}\right)^2 d\xi + pv_2$

Since the beam is fixed at $\xi = 0 \Rightarrow v_1 = \theta_1 = 0$ can be replaced in v(ξ) to give:

$\Pi = \dfrac{EI}{2l^3}\left(12v_2^{\,2} + 2\theta_2^{\,2}l^2 - 12v_2\theta_2 l\right) + pv_2$ and defining: $f = \dfrac{2\Pi l^3}{EI}$; $x_1 = v_2$; $x_2 = \theta_2 l \Rightarrow$

$$\boxed{f = 12x_1^{\,2} + 4x_2^{\,2} - 12x_1x_2 + 2x_1}$$

Example

$$\boxed{\text{Min} \quad f = 12x_1^2 - 12x_1x_2 + 4x_2^2 + 2x_1}$$

$\mathbf{x}_0^1 = (-1,-2)^T$ e $f(\mathbf{x}_0^1) = 2$
$\left(\mathbf{x}^* = (-1/3,-1/2)^T \text{ exact solution}\right)$

Solution:

$\mathbf{s}_1^1 = (1,0)^T \Rightarrow \mathbf{x}_1^1 = \begin{Bmatrix} -1 \\ -2 \end{Bmatrix} + \alpha \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -1+\alpha \\ -2 \end{Bmatrix} \Rightarrow f(\alpha) = 12(-1+\alpha)^2 + 4(-2)^2 - 12(-1+\alpha)(-2) +$
$+2(-1+\alpha) \Rightarrow \text{Min f} \Rightarrow \alpha = -1/12 \Rightarrow \mathbf{x}_1^1 = \begin{Bmatrix} -13/12 \\ -2 \end{Bmatrix}$ and $f(\mathbf{x}_1^1) = 1{,}9166$; $\mathbf{s}_2^1 = (0,1)^T \Rightarrow \mathbf{x}_2^1 = \begin{Bmatrix} -13/12 \\ -2 \end{Bmatrix} +$
$+\alpha \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -13/12 \\ -2+\alpha \end{Bmatrix} \Rightarrow \text{Min f} \Rightarrow \alpha = 3/8 \Rightarrow \mathbf{x}_2^1 = \begin{Bmatrix} -13/12 \\ -13/8 \end{Bmatrix}$ and $f(\mathbf{x}_2^1) = 1{,}3541 \Rightarrow \mathbf{s}_p^1 = \mathbf{x}_2^1 - \mathbf{x}_0^1 = \begin{Bmatrix} -1/12 \\ 3/8 \end{Bmatrix}$
$\Rightarrow$
$\Rightarrow \mathbf{x}_0^2 = \begin{Bmatrix} -1 \\ -2 \end{Bmatrix} + \alpha \begin{Bmatrix} -1/12 \\ 3/8 \end{Bmatrix} = \begin{Bmatrix} -1-\alpha/12 \\ -2+3\alpha/8 \end{Bmatrix} \Rightarrow \text{Min f} \Rightarrow \alpha = 40/49 \Rightarrow \mathbf{x}_0^2 = \begin{Bmatrix} -157/147 \\ -83/49 \end{Bmatrix}$ and $f(\mathbf{x}_0^2) = 1{,}31972$

$$|\alpha| = \frac{40}{49} < \left[\frac{2 - 1{,}31972}{1{,}9166 - 1{,}3541}\right]^{\frac{1}{2}}$$

Cycle	x_1	x_2	f
0	−1.0	−2.0	2.0
1	−1.083334	−2.0	1.916667
1	−1.083334	−1.625	1.354167
2	−0.895834	−1.625	0.9322967
2	−0.895834	−1.34375	0.6158854
2	−0.33334	−0.499999	−0.333333

First-Order Methods

Employ objective function values and its derivatives (gradient) relative to the design variables.

First-order methods present linear convergence rate or superlinear, i.e., considering:

$$\left\| \mathbf{x}_{k+1} - \mathbf{x}^* \right\| \le c_k \left\| \mathbf{x}_k - \mathbf{x}^* \right\|^p$$

the sequence $\mathbf{x}_k$ presents:

- linear convergence for $\mathbf{x}^*$, if p=1 and constant c_k.
- superlinear convergence for $\mathbf{x}^*$, if p=1 and c_k is a sequence that converges to zero.

First-Order Methods

Steepest Descent Method

$$\min_{\mathbf{s}} \quad \nabla f^T \mathbf{s} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} s_i$$

Subject to $\mathbf{s}^T\mathbf{s} = 1$

$$L(\mathbf{s}, \lambda) = \nabla f^T \mathbf{s} + \lambda(\mathbf{s}^T\mathbf{s} - 1) \Rightarrow \frac{\partial L}{\partial \mathbf{s}} = \nabla f + 2\lambda \mathbf{s} = 0 \Rightarrow$$

$$\Rightarrow \mathbf{s} = -\frac{\nabla f}{2\lambda} \Rightarrow \mathbf{s}^T\mathbf{s} = 1 \Rightarrow \|\nabla f\|^2 = 4\lambda^2 \Rightarrow 2\lambda = \|\nabla f\| \Rightarrow$$

Solution: $\boxed{\mathbf{s} = -\frac{\nabla f}{\|\nabla f\|}}$ where || || is the Eucledian norm

Thus: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{s}$ ➡ one-dimensional (search) optimization

If the function f is quadratic: $f = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \Rightarrow \alpha^* = -\frac{(\mathbf{x}_k{}^T\mathbf{Q} + \mathbf{b}^T)\mathbf{s}}{\mathbf{s}^T\mathbf{Q}\mathbf{s}}$

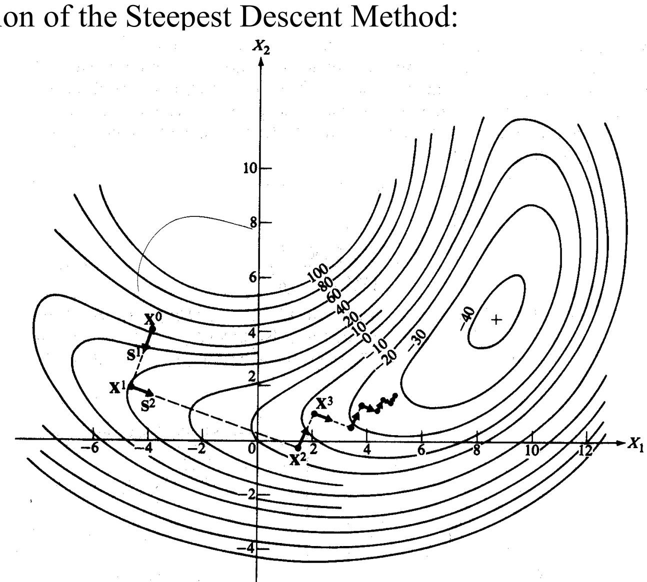
symmetric

Method's Performance ➡ Conditioning Number (CN) of the matrix **Q**:

$CN = \frac{\lambda_{max}}{\lambda_{min}}$ if CN is large ➡ slow progress and "zig-zag" pattern ("hemstitching" phenomenon)

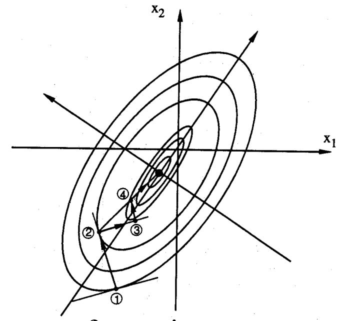
First-Order Methods

Interpretation of the Steepest Descent Method:



First-Order Methods

Now, let us see the example: $f = 12x_1^2 - 12x_1x_2 + 4x_2^2 + 2x_1$

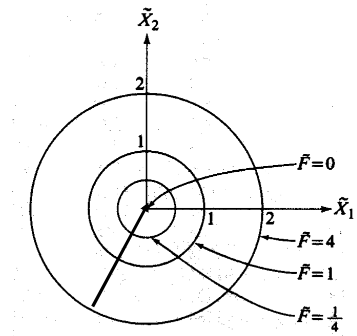


But, considering the transformation of variables:

$y_1 = \left(x_1 - \frac{1}{2}x_2\right);\; y_2 = \frac{x_2}{\sqrt{12}} \Rightarrow$

$\Rightarrow f(y_1, y_2) = y_1^2 + y_2^2 + \frac{1}{6}\left(y_1 + \sqrt{3}y_2\right)$

$(CN = 1)$



First-Order Methods

Thus, the method presents a linear rate of convergence without the change of variables. But because it is difficult to find the appropriate transformation ➡ Conjugate Gradient Method

Conjugate Gradients Method ("Fletcher-Reeves")

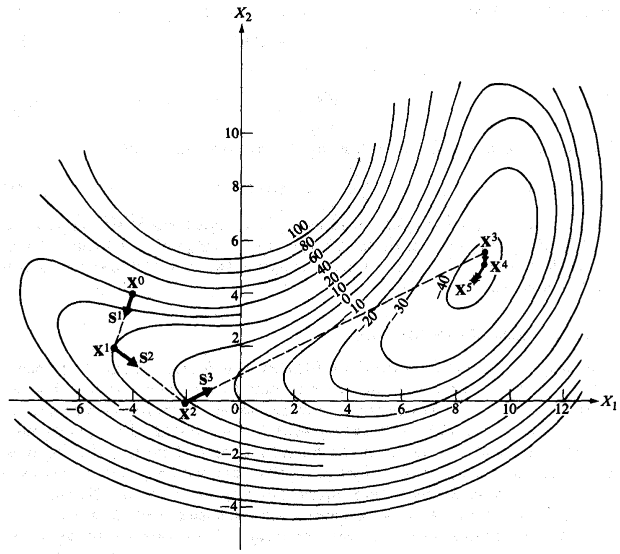
Directions $\mathbf{s}_{k+1}$ and $\mathbf{s}_k$ are Q-conjugate: $\Rightarrow \mathbf{s}^T{}_{k-1}\mathbf{Q}\mathbf{s}_k = 0$

Algorithm:

1. Compute $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_{k+1}\mathbf{s}_k$ where α_{k+1} is determined such that $\frac{df(\alpha_{k+1})}{d\alpha_{k+1}} = 0$
2. $\mathbf{s}_k = \mathbf{g}_k = -\nabla f(\mathbf{x}_k)$ if $k = 0$ and $\mathbf{s}_k = \mathbf{g}_k + \beta_k \mathbf{s}_{k-1}$ if $k > 0$ with
$$\beta_k = \frac{\mathbf{g}_k{}^T\mathbf{g}_k}{\mathbf{g}_{k-1}{}^T\mathbf{g}_{k-1}} \quad \text{e} \;\; \mathbf{g}_k = -\nabla f(\mathbf{x}_k)$$
3. If $\|\mathbf{g}_{k+1}\|$ or $|f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)|$ is sufficiently small, stop. Otherwise:
4. If k<n go to 1, or restart.

First-Order Methods

Interpretation of the Conjugate Gradient Method:



Example

$$\boxed{\text{Min} \quad f = 12x_1^2 - 12x_1x_2 + 4x_2^2 + 2x_1}$$

Solution:

$$\mathbf{x}_0{}^T = (-1,-2) \Rightarrow \nabla f(\mathbf{x}_0) = \begin{Bmatrix} 24x_1 - 12x_2 + 2 \\ 8x_2 - 12x_1 \end{Bmatrix}_{\mathbf{x}=\mathbf{x}_0} \Rightarrow \mathbf{s}_0 = -\nabla f(\mathbf{x}_0) = \begin{Bmatrix} -2 \\ 4 \end{Bmatrix} \Rightarrow \mathbf{x}_1$$

$$= \begin{Bmatrix} -1 \\ -2 \end{Bmatrix} + \alpha_1 \begin{Bmatrix} -2 \\ 4 \end{Bmatrix} \Rightarrow$$

$$\Rightarrow f(\alpha_1) = 12(-1-2\alpha_1)^2 + 4(-1+4\alpha_1{}^2)^2 - 12(-1-2\alpha_1)(-2+4\alpha_1) + 2(-1-2\alpha_1)$$

$$\Rightarrow \frac{df}{d\alpha_1} = 0 \Rightarrow$$

$$\Rightarrow \alpha_1 = 0{,}048077 \Rightarrow \mathbf{x}_1 = \begin{Bmatrix} -1{,}0961 \\ -1{,}8077 \end{Bmatrix} \text{ and } \nabla f(\mathbf{x}_1) = \begin{Bmatrix} -2{,}6154 \\ -1{,}3077 \end{Bmatrix} \Rightarrow \mathbf{s}_1 = -\nabla f(\mathbf{x}_1) + \beta_1 \mathbf{s}_0 \Rightarrow$$

$$\Rightarrow \beta_1 = \frac{(-2.6154)^2 + (-1.3077)^2}{(-2)^2 + (4)^2} = 0{,}4275 \Rightarrow \mathbf{s}_1 = \begin{Bmatrix} -2{,}6154 \\ -1{,}3077 \end{Bmatrix} + 0{,}4275 \begin{Bmatrix} -2 \\ 4 \end{Bmatrix}$$

$$= \begin{Bmatrix} 1{,}76036 \\ 3{,}0178 \end{Bmatrix} \text{ and}$$

$$\mathbf{x}_2 = \begin{Bmatrix} -1{,}0961 \\ -1{,}8077 \end{Bmatrix} + \alpha_2 \begin{Bmatrix} 1{,}76036 \\ 3{,}0178 \end{Bmatrix} \Rightarrow \frac{df(\alpha_2)}{d\alpha_2} = 0 \Rightarrow \alpha_2 = 0.4334 \Rightarrow \mathbf{x}_2 = \mathbf{x}^* = \begin{Bmatrix} -0{,}3334 \\ -0{,}5 \end{Bmatrix}$$

$$\text{and } \nabla f(\mathbf{x}_2) = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \Rightarrow \begin{Bmatrix} -2 \\ 4 \end{Bmatrix} \begin{bmatrix} 24 & -12 \\ -12 & 8 \end{bmatrix} \begin{Bmatrix} 1{,}76036 \\ 3{,}0178 \end{Bmatrix} \cong 0$$

Second Order Methods

Newton's Method

The direction **s** is obtained through the solution of the following problem:

$$\begin{array}{ll} \underset{\mathbf{s}}{\text{Min}} & \nabla f^T \mathbf{s} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} s_i \\ \text{tal que} & \mathbf{s}^T \mathbf{Q} \mathbf{s} = 1 \end{array}$$

Solution: $L(\mathbf{s},\lambda) = \nabla f^T \mathbf{s} + \lambda(\mathbf{s}^T \mathbf{Q}\mathbf{s} - 1) \Rightarrow \frac{\partial L}{\partial \mathbf{s}} = \nabla f + 2\lambda \mathbf{Q}\mathbf{s} = 0 \Rightarrow$

$$\Rightarrow \mathbf{s} = -\frac{\mathbf{Q}^{-1}\nabla f}{2\lambda} \Rightarrow \mathbf{s} = -\mathbf{Q}^{-1}\nabla f \Rightarrow \mathbf{Q}\mathbf{s} = -\nabla f$$

Must be positive definite

$$\Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_{k+1}\mathbf{s}_k$$

$\mathbf{Q}\mathbf{s} = -\nabla f$
- n^3 operations in the solution
- $\frac{n(n+1)}{2}$ elements must be calculated to assemble Q

➡ Quasi-Newton Method

Second Order Methods

Quasi-Newton Method

$$\underbrace{\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)}_{\mathbf{y}_k} \cong \underbrace{\mathbf{Q}}_{\mathbf{A}_k}\underbrace{(\mathbf{x}_{k+1} - \mathbf{x}_k)}_{\mathbf{p}_k} \Rightarrow \mathbf{y}_k = \mathbf{A}_k\mathbf{p}_k \text{ e } \boxed{\mathbf{B}_{k+1}\mathbf{y}_k = \mathbf{p}_k}$$

approximation

Seccant relation or Quasi-Newton

$\mathbf{B}_{k+1}$ is the approximation of $\mathbf{Q}^{-1}$. Eventually: $\mathbf{B}_{k+1}\mathbf{A}_k = \mathbf{I}$

$\mathbf{A}_k$ and $\mathbf{B}_k$ must stay positive definite. After obtaining $\mathbf{B}_k$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_{k+1}\mathbf{s}_k \text{ and } \mathbf{s}_k = -\mathbf{B}_k\nabla f(\mathbf{x}_k)$$

BFGS approximation for $\mathbf{B}_k$: $\mathbf{B}_{k+1} = \left[\mathbf{I} - \frac{\mathbf{p}_k\mathbf{y}_k^T}{\mathbf{p}_k^T\mathbf{y}_k}\right]\mathbf{B}_k\left[\mathbf{I} - \frac{\mathbf{y}_k^T\mathbf{p}_k}{\mathbf{p}_k^T\mathbf{y}_k}\right] + \frac{\mathbf{p}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{y}_k}$

Example

$$\boxed{\text{Min} \quad f = 12x_1^2 - 12x_1x_2 + 4x_2^2 + 2x_1} \quad \mathbf{x}_0^T = (-1,-2);$$ Solution:

We start with the direction of the "Steepest Descent" Method and $\mathbf{B}_0$=**I** (positive definite):

$$\mathbf{x}_1 = \begin{Bmatrix} -1.0961 \\ -1.8077 \end{Bmatrix} \text{ and } \nabla f(\mathbf{x}_1) = \begin{Bmatrix} -2.6154 \\ -1.3077 \end{Bmatrix} \Rightarrow \mathbf{p}_0 = \begin{Bmatrix} -1.0961 \\ -1.8077 \end{Bmatrix} - \begin{Bmatrix} -1 \\ -2 \end{Bmatrix} = \begin{Bmatrix} -0.0961 \\ 0.1923 \end{Bmatrix} \text{ and because}$$

$$\mathbf{s}_0 = -\mathbf{B}_0 \nabla f(\mathbf{x}_0) = -\mathbf{I} \nabla f(\mathbf{x}_0) = \begin{Bmatrix} 2 \\ -4 \end{Bmatrix} \Rightarrow$$

$$\mathbf{y}_0 = \begin{Bmatrix} -2.6154 \\ -1.3077 \end{Bmatrix} - \begin{Bmatrix} 2 \\ -4 \end{Bmatrix} = \begin{Bmatrix} -4.6154 \\ 2.6923 \end{Bmatrix} \Rightarrow \mathbf{p}_0 \mathbf{y}_0^T = 0.96127 \text{ and}$$

$$\mathbf{p}_0^T \mathbf{y}_0 = \begin{Bmatrix} -0.0961 \\ 0.1923 \end{Bmatrix} \{-4.6154 \quad 2.6923\} = \begin{bmatrix} 0.44354 & -0.25873 \\ -0.88754 & 0.51773 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \mathbf{B}_1$$

$$= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{0.96127} \begin{bmatrix} 0.44354 & -0.25873 \\ -0.88754 & 0.51773 \end{bmatrix} \right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{0.96127} \begin{bmatrix} 0.44354 & -0.88754 \\ -0.25873 & 0.51773 \end{bmatrix} \right) +$$

$$+ \frac{1}{0.96127} \begin{bmatrix} 0.00923 & -0.01848 \\ -0.01848 & 0.03698 \end{bmatrix} = \begin{bmatrix} 0.37213 & 0.60225 \\ 0.60225 & 1.10385 \end{bmatrix} \Rightarrow$$

$$\mathbf{s}_1 = -\begin{bmatrix} 0.37213 & 0.60225 \\ 0.60225 & 1.10385 \end{bmatrix} \begin{Bmatrix} -2.6154 \\ -1.3077 \end{Bmatrix} = \begin{Bmatrix} 1.7608 \\ 3.0186 \end{Bmatrix} \Rightarrow$$

$$\mathbf{x}_2 = \begin{Bmatrix} -1.0961 \\ -1.8077 \end{Bmatrix} + \alpha_2 \begin{Bmatrix} 1.7608 \\ 3.0186 \end{Bmatrix} \Rightarrow \mathbf{x}_2 = \begin{Bmatrix} -0.3333 \\ -0.5 \end{Bmatrix} \text{ and } \nabla f(\mathbf{x}_2) \cong \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \text{ (convergence)}$$

And it can be verified that: $\mathbf{B}_2 = \begin{bmatrix} 0.1667 & 0.25 \\ 0.25 & 0.5 \end{bmatrix} = \mathbf{Q}^{-1}$

SCILAB implementation

SCILAB routine to solve unconstrained problems: **OPTIM**

[f,[xopt,[gradopt,[work]]]]=optim(costf,[contr],x0,['sth'],[df0,[mem]], [work],[stop],['in']) solves the problem:

$$\boxed{\begin{array}{l} \text{Min}_{\mathbf{x}} \ \text{costf}(\mathbf{x}) \\ \text{Subject to } \mathbf{b}_i \le \mathbf{x} \le \mathbf{b}_s \end{array}}$$

- costf – function to be minimized
- x_{opt} – vector for the optimal solution;
- x_0 – start guess of x;
- f – optimal value of the function;
- contr - 'b', $\mathbf{b}_i$, $\mathbf{b}_s$;
- sth - 'qn' (Quasi-Newton); 'gc' (conjugate gradients); 'nd' (non-differentiable – zeroth order method);

SCILAB implementation

• mem – number of variables used to approximate the Hessian matrix;
• stop - 'ar' (keyword); 'nap' (maximum number of computations of f); 'iter' (maximum number of allowed iterations); 'epsg' (cutoff in the value of gradient norm); 'epsf' (cutoff in the variation of costf); 'epsx' (vector with cutoff values for **x**);
• gradopt - costf gradient is supplied;
• work – vector for restart;

Unconstrained Nonlinear Optimization Algorithms

Further considerations:

•Solution of Linear and Nonlinear Sets of Equations:

Necessary Condition for the Minimization of the Potential Energy ➡ Gradient of the energy is null ➡ $\nabla f(\mathbf{x}) = g(\mathbf{x}) = 0$

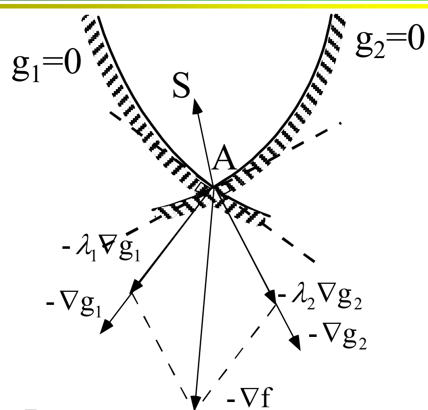
Problem equivalent to: Min $f = \frac{1}{2} g^T g$ i.e.:

$$\mathbf{Ax} = \mathbf{b} \ \equiv \ \text{Min } \frac{1}{2}(\mathbf{Ax}-\mathbf{b})^T(\mathbf{Ax}-\mathbf{b})$$

In the nonlinear structural analysis are obtained the stable and unstable equilibrium positions (if the method does not converge to a local minimum)

• Probabilistic Algorithms:
 * search for the global minimum
 * powerful tools for problems with discrete variables
 * use the process of random search guided by probabilistic decisions.

KKT Optimality Conditions



$\mathbf{s}^T\nabla g_j \geq 0$ (viable direction)

$\mathbf{s}^T\nabla f < 0$ (useful direction)

Since $\mathbf{s}^T\nabla f = \sum_{j=1}^{ng} \lambda_j \mathbf{s}^T\nabla g_j \Rightarrow$

$\Rightarrow$ $\mathbf{s}^T\nabla f < 0$ is impossible if $\lambda_j \geq 0$

$\mathbf{x}^*$ is a local minimum:

$$L(\mathbf{x},\lambda) = f(\mathbf{x}) - \sum_{j=1}^{n_g} \lambda_j (g_j - t_j^2) \Rightarrow$$

1. $\mathbf{x}^*$ is viable $\Rightarrow$ $g_j(\mathbf{x}^*) \leq 0$ $j = 1, n_g$
2. $\nabla f(\mathbf{x}^*) - \sum_{j=1}^{n_g} \lambda^*{}_j \nabla g_j(\mathbf{x}^*) = 0;$
   $\lambda^*{}_j \geq 0;$
3. $\lambda^*{}_j g_j(\mathbf{x}^*) = 0 \Rightarrow \therefore$ se $\lambda_j = 0 \Rightarrow$
   $\Rightarrow$ g_j is not active

Particular case:

$$\underbrace{\mathbf{s}^T\nabla f}_{\perp \text{ gradient}} = \underbrace{\mathbf{s}^T\nabla g_j}_{\textbf{Tangent to constraint}} = 0 \Rightarrow \text{higher derivatives}$$

KKT Optimality Conditions

KKT are necessary conditions, but are not sufficient for optimality

KKT are sufficient if:

- $n_{\text{active constraints}} = n_{\text{variables}}$:

$$\underbrace{\mathbf{s}^T \overbrace{\nabla g_j}^{\textbf{n linearly independent directions}}}_{\textbf{n linearly independent equations}} = 0 \quad \text{only solution: } \mathbf{s} = 0$$

Obs.: active constraints with linearly independent gradients

- $n_{\text{active constraints}} \neq n_{\text{variables}}$: sufficiency condition requires higher order derivatives ➡ Hessian matrix must be positive definite:

$\mathbf{s}^T(\nabla^2 L)\mathbf{s} > 0$ for any $\mathbf{s}$ such that $\mathbf{s}^T\nabla g_j = 0$ (active constraints with $\lambda_j > 0$)

- Convex Problems

Example

Min $f = -x_1^3 - 2x_2^2 + 10x_1 - 6 - 2x_2^3$
x_1, x_2
Subject to

$$g_1 = 10 - x_1 x_2 \geq 0$$
$$g_2 = x_1 \geq 0$$
$$g_3 = 10 - x_2 \geq 0$$

Solution: KKT conditions:

$$L = -x_1^3 - 2x_2^2 + 10x_1 - 6 - 2x_2^3 - \lambda_1(10 - x_1x_2) - \lambda_2 x_1 + - \lambda_3(10 - x_2)$$

$$\frac{\partial f}{\partial x_1} + \lambda_1 \frac{\partial g_1}{\partial x_1} + \lambda_2 \frac{\partial g_2}{\partial x_1} = -3x_1^2 + 10 + \lambda_1 x_2 - \lambda_2 = 0$$

$$\frac{\partial f}{\partial x_2} + \lambda_1 \frac{\partial g_1}{\partial x_2} + \lambda_3 \frac{\partial g_3}{\partial x_2} = -4x_2 - 6x_2^2 + \lambda_1 x_1 - \lambda_3 = 0$$

Possible cases to be analyzed:

	λ_1	λ_2	λ_3
Case 1	0	0	0
Case 2	$\neq 0$	0	$\neq 0$
Case 3	$\neq 0$	0	0
Case 4	0	$\neq 0$	0
Case 5	0	0	$\neq 0$
Case 6	0	$\neq 0$	$\neq 0$
Case 7	$\neq 0$	$\neq 0$	$\neq 0$

Case 7 → Not possible

$n_{\text{active constraints}} \neq n_{\text{variables}}$?

Hessian matrix:

$$\frac{\partial^2 L}{\partial x_1^2} = -6x_1; \quad \frac{\partial^2 L}{\partial x_2^2} = -4 - 12x_2; \quad \frac{\partial^2 L}{\partial x_1 \partial x_2} = -\lambda_1 = \frac{\partial^2 L}{\partial x_2 \partial x_1} \Rightarrow \nabla^2 L = \begin{bmatrix} -6x_1 & -\lambda_1 \\ -\lambda_1 & -4 - 12x_2 \end{bmatrix}$$

Example

Case 1: $x_1 = 1{,}826; x_2 = 0 \Rightarrow f = 6{,}17 \Rightarrow \nabla^2 L = \begin{bmatrix} -6x_1 & \lambda_1 \\ \lambda_1 & -4 - 12x_2 \end{bmatrix}$ **Negative definite** ➡ **Maximal Point**

Case 2: $x_1 = 1; x_2 = 10; \lambda_1 = \text{-}0{,}7; \lambda_3 = 639{,}3$ **Not a minimum, nor a maximum**

Case 3: $\begin{cases} -3x_1^2 + 10 + \lambda_1 x_2 = 0 \\ -4x_2 - 6x_2^2 + \lambda_1 x_1 = 0 \\ x_1 x_2 = 10 \end{cases} \Rightarrow \begin{array}{l} x_1 = 3{,}847 \\ x_2 = 2{,}599 \\ \lambda_1 = 13{,}24 \\ f = -73{,}08 \end{array} \Rightarrow \nabla^2 L = \begin{bmatrix} -23{,}08 & 13{,}24 \\ 13{,}24 & -35{,}19 \end{bmatrix}$ **Negative definite**

Case 4: $\begin{cases} -3x_1^2 + 10 - \lambda_2 = 0 \\ -4x_2 - 6x_2^2 = 0 \end{cases} \Rightarrow \begin{array}{llll} x_1 = 0 & x_2 = 0 & \lambda_2 = 10 & f = -6 \\ x_1 = 0 & x_2 = -2/3 & \lambda_2 = 10 & f = -6{,}99 \end{array}$ $\Rightarrow \mathbf{s}^T \nabla^2 L \mathbf{s} < 0$ **Not a minimum**

Case 5: $\begin{cases} -3x_1^2 + 10 = 0 \\ -4x_2 - 6x_2^2 - \lambda_3 = 0 \end{cases} \Rightarrow x_1 = 1{,}826 \quad x_2 = 10 \quad \lambda_3 = 640 \quad f = -2194 \Rightarrow \nabla^2 L$ **Negative definite**

Case 6: $x_1 = 0; \; x_2 = 10; \; \lambda_2 = 10; \; \lambda_3 = 640; \; f = \text{-}2206$ **KKT is satisfied and the number of active constraints = number of design variables**

Calculation of Lagrange Multipliers

Important for verification of the KKT conditions of the optimal solution

Let the condition of optimality be: $\nabla f - \mathbf{N}\lambda = 0$

where:

$$n_{ij} = \frac{\partial g_j}{\partial x_i} \quad j = 1,\dots,r \quad \text{(active constraints)}$$

$$i = 1,\dots,n \quad \text{(number of design variables)}$$

$r < n \Rightarrow$ number of unknowns $(\lambda) <$ number of equations

Solution: Least squares

$$u = \mathbf{N}\lambda - \nabla f \Rightarrow \text{Min } \|u\|^2 = (\mathbf{N}\lambda - \nabla f)^T(\mathbf{N}\lambda - \nabla f) \Rightarrow$$

$$\Rightarrow 2\mathbf{N}^T\mathbf{N}\lambda - 2\mathbf{N}^T\nabla f = 0 \Rightarrow \boxed{\lambda = (\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T\nabla f}$$

If the stationary condition is satisfied, the solution is exact, therefore:

$$\mathbf{P}\nabla f = 0, \qquad \text{where } \mathbf{P} = \mathbf{I} - \mathbf{N}(\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T$$

$\mathbf{P}$ - Projection Matrix

Example

Min $\quad f = x_1 + x_2 + x_3$

x_1,x_2,x_3

Subject to $g_1 = 8 - x_1^2 - x_2^2 \geq 0$

$g_2 = x_3 - 4 \geq 0$

$g_3 = x_2 + 8 \geq 0$

Let us check if (-2,-2,4) is a local minimum:

$g_1 = 0;\ g_2 = 0;\ g_3 = 6 \Rightarrow$

$$\left.\begin{matrix} \dfrac{\partial g_1}{\partial x_1} = -2x_1 = 4;\ \dfrac{\partial g_1}{\partial x_2} = -2x_2 = 4;\ \dfrac{\partial g_1}{\partial x_3} = 0 \\ \dfrac{\partial g_2}{\partial x_1} = 0;\ \dfrac{\partial g_2}{\partial x_2} = 0;\ \dfrac{\partial g_2}{\partial x_3} = 1 \end{matrix}\right\} \Rightarrow \mathbf{N} = \begin{bmatrix} 4 & 0 \\ 4 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial x_3} = 1 \Rightarrow \nabla f = \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix};\ \mathbf{N}^T\mathbf{N} = \begin{bmatrix} 32 & 0 \\ 0 & 1 \end{bmatrix};\ \mathbf{N}^T\nabla f = \begin{Bmatrix} 8 \\ 1 \end{Bmatrix} \Rightarrow \lambda = (\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T\nabla f = \begin{Bmatrix} 1/4 \\ 4 \end{Bmatrix}$$

$$\mathbf{P}\nabla f = \left[\mathbf{I} - \mathbf{N}(\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T\right]\nabla f = 0$$

And all Lagrange multipliers are positive, but, since $r \neq n$ ➡ Hessian matrix must be obtained to verify KKT

Note that:

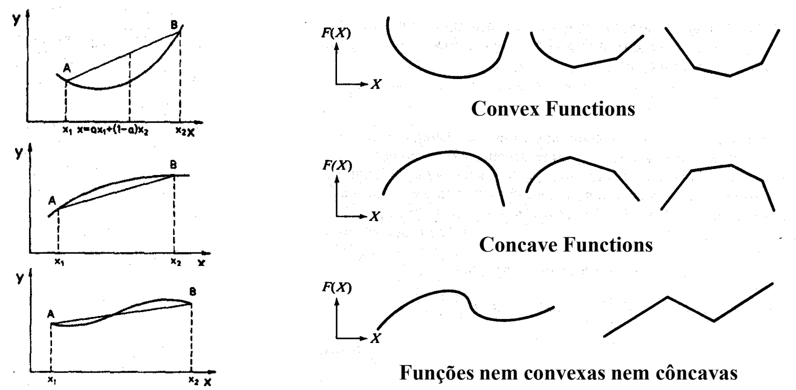
$$\nabla f - \mathbf{N}\lambda = 0 \Rightarrow \begin{cases} \left.\begin{matrix} 1 - 4\lambda_1 = 0 \\ 1 - 4\lambda_1 = 0 \end{matrix}\right\} \textbf{Linearly dependent} \\ 1 - \lambda_2 = 0 \end{cases}$$

Convex Problems

Convex problem ➡ Objective function and feasible domain are complex

Convex Function

$$f[\alpha \mathbf{x}_2 + (1-\alpha)\mathbf{x}_1] \le \alpha f(\mathbf{x}_2) + (1-\alpha) f(\mathbf{x}_1), \quad 0 < \alpha < 1$$ Examples:



Convex Function ➡ Hessian matrix is positive semi-definite;

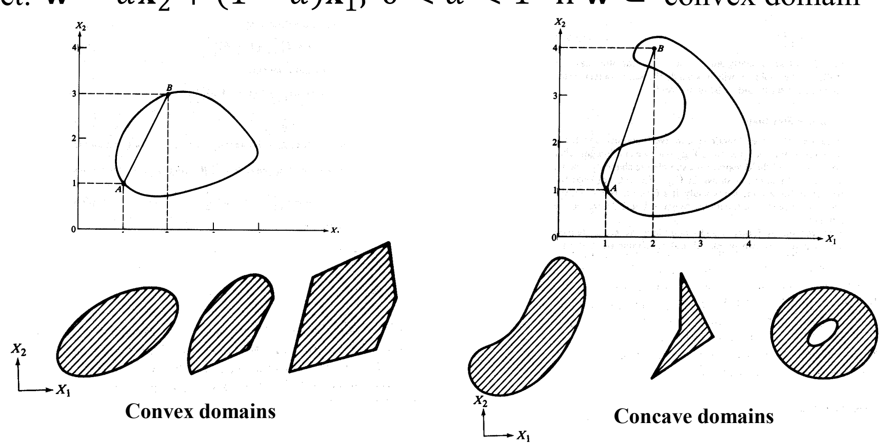
Concave Function ➡ Hessian matrix is negative semi-definite;

A linear function is convex.

Convex Problems

Convex Domain

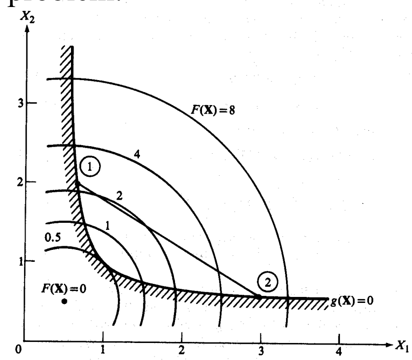
Let: $\mathbf{w} = \alpha \mathbf{x}_2 + (1-\alpha)\mathbf{x}_1,\ 0 < \alpha < 1$ If $\mathbf{w} \subset$ convex domain



Convex domain if $\begin{cases} \text{All inequality constraints are concave} \\ \text{All equality constraints are linear} \end{cases}$

Convex Problems

Example of convex problem:



Advantage of convex problems:

- Global optimum
- KKT also become sufficient conditions

The major part of the problems are nonconvex ➡ Approximated by a series of convex problems, for example, linear programming.

Example

Verify if the problem is convex:

$$\min_{x_1, x_2} \quad f = 3x_1 + \sqrt{3}x_2$$

Subject to

$$g_1 = 3 - \frac{18}{x_1} - \frac{6\sqrt{3}}{x_2} \geq 0$$

$$g_2 = x_1 - 5{,}73 \geq 0$$

$$g_3 = x_2 - 7{,}17 \geq 0$$

$f, g_2, g_3 \Rightarrow$ Linear functions

Function g_1: Hessian value given by $\mathbf{A}_1 = \begin{bmatrix} -36/x_1^3 & 0 \\ 0 & -12\sqrt{3}/x_2^3 \end{bmatrix}$

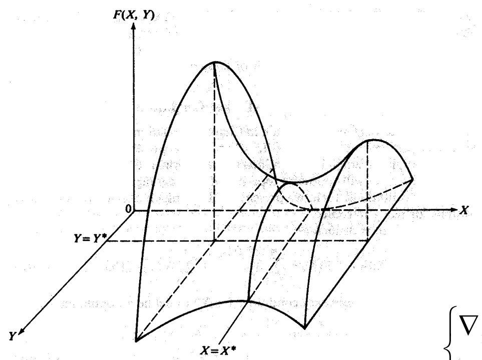
Thus the Hessian matrix is negative definite ➡ g_1 is concave

Therefore, the objective function is linear (convex) and the constraints are linear or concave (convex domain) ➡ the problem is convex.

Thus, KKT are sufficient conditions for the minimum.

Ponto Sela (“Saddle Point”)

31



Uma função F(**x**,**y**) possue ponto “sela” se:

$$F(\mathbf{x}^*,\mathbf{y}) \le F(\mathbf{x}^*,\mathbf{y}^*) \le F(\mathbf{x},\mathbf{y}^*)$$

$$F(\mathbf{x}^*,\mathbf{y}^*)\begin{cases}\text{mínimo em } \mathbf{x}\\ \text{máximo em } \mathbf{y}\end{cases}$$

No caso da função Lagrangeana: $L(\mathbf{x}^*,\lambda^*)$

$$\begin{cases}\nabla_x L(\mathbf{x}^*,\lambda^*)=0\\ \nabla_\lambda L(\mathbf{x}^*,\lambda^*)=0\end{cases} \quad \text{e} \quad \begin{cases}\nabla_\lambda L(\mathbf{x}^*,\lambda)\le 0\\ \nabla_x L(\mathbf{x},\lambda^*)\ge 0\end{cases}$$

Solução do Ponto Sela ➡ Problema Min Max:

$$L(\lambda)=\min_{\mathbf{x}} L(\mathbf{x},\lambda) \Rightarrow \max_{\lambda} L(\lambda) = \max_{\lambda}\min_{\mathbf{x}} L(\mathbf{x},\lambda)$$

ou:

$$L(\mathbf{x})=\max_{\lambda} L(\mathbf{x},\lambda) \Rightarrow \min_{\mathbf{x}} L(\mathbf{x}) = \min_{\mathbf{x}}\max_{\lambda} L(\mathbf{x},\lambda)$$

$$\boxed{\max_{\lambda}\min_{\mathbf{x}} L(\mathbf{x},\lambda) \equiv \min_{\mathbf{x}}\max_{\lambda} L(\mathbf{x},\lambda)}$$

Prof. Dr. Emilio C. Nelli Silva

Exemplo

32

Ex.:

$$\min_{x} \frac{1}{x}$$
tal que $x-1 \le 0$
$x \ge 0$

O Lagrangeano do problema vale:

$$L(x,\lambda)=\frac{1}{x}+\lambda(x-1)$$

Calculando o mínimo em x: $\nabla_x L(x,\lambda)=0 \Rightarrow -\frac{1}{x^2}+\lambda=0 \Rightarrow$

$$\Rightarrow x=\pm\frac{1}{\sqrt{\lambda}}\ (\text{mas } x\ge 0) \Rightarrow x=\frac{1}{\sqrt{\lambda}} \Rightarrow L(\lambda)=\sqrt{\lambda}+\lambda\left(\frac{1}{\sqrt{\lambda}}-1\right)=2\sqrt{\lambda}-\lambda$$

Calculando o máximo em λ: $\nabla_\lambda L(\lambda)=\frac{1}{\sqrt{\lambda}}-1 \Rightarrow \lambda^*=1 \Rightarrow x^*=1$

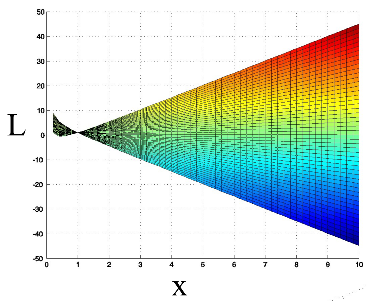
Ponto “Sela”: $\boxed{\therefore (x^*,\lambda^*)=(1,1)}$

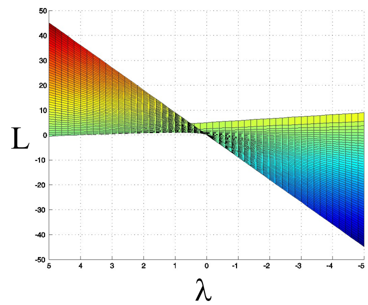
Os gradientes devem ser contínuos em **x** e λ , caso contrário, outra abordagem de solução para o problema Min Max deve ser utilizada.

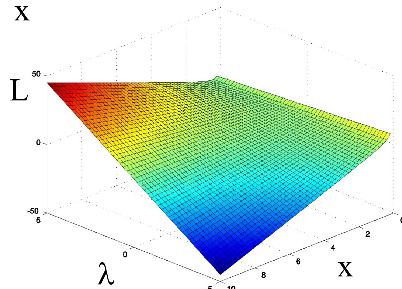
Prof. Dr. Emilio C. Nelli Silva

Example

Plotagem da função Lagrangeana:







Duality

Primal problem		Dual problem
$\underset{\mathbf{x}}{\text{Min}} \quad f_p = \mathbf{c}^T\mathbf{x}$ Subject to $\mathbf{Ax} \leq \mathbf{b}$ $\mathbf{x} \geq \mathbf{0}$	$\equiv$	$\underset{\lambda}{\text{Max}} \quad f_D = -\lambda^T\mathbf{b}$ Subject to $\mathbf{A}^T\lambda \geq -\mathbf{c}$ $\lambda \geq \mathbf{0}$
m constraints n design variables m>n	$\boxed{f_{p\,min} = f_{D\,max}}$	n constraints m design variables m>n

Proof:

x* is minimal

$$L(\mathbf{x},\lambda) = \mathbf{c}^T\mathbf{x} + \lambda^T\{\mathbf{Ax} - \mathbf{b}\} \Rightarrow \nabla_{\mathbf{x}} L(\mathbf{x},\lambda) = \mathbf{c} + \mathbf{A}^T\lambda \boxed{\geq 0} \Rightarrow \mathbf{cx}^* + \lambda^T\mathbf{Ax}^* = 0 \Rightarrow$$

$L(\lambda) = -\lambda^T\mathbf{b}$ must be maximized ➡ Dual problem and: $f(\mathbf{x}^*) = L(\lambda^*)$

If: $\lambda_j \neq 0 \Rightarrow j^{th}$ constraint is active in primal problem
$\lambda_j = 0 \Rightarrow j^{th}$ constraint is inactive in primal problem

Example

Min $f = -4x_1 - x_2 + 50$
x_1,x_2
Subject to
$x_1 - x_2 \le 2$
$x_1 + 2x_2 \le 8$
$x_1 + x_2 \le 10$
$-5x_1 + x_2 \le 5$
$x_1 \ge 0;\ \ x_2 \ge 0$

$$\Rightarrow \mathbf{c} = \begin{Bmatrix} -4 \\ -1 \end{Bmatrix}; \mathbf{A} = \begin{Bmatrix} 1 & -1 \\ 1 & 2 \\ 1 & 1 \\ -5 & 1 \end{Bmatrix}; \mathbf{b} = \begin{Bmatrix} 2 \\ 8 \\ 10 \\ 5 \end{Bmatrix} \Rightarrow$$

Max $L = -2\lambda_1 - 8\lambda_2 - 10\lambda_3 - 5\lambda_4 + 50$
$\lambda_1, \lambda_2, \lambda_3, \lambda_4$
Subject to
$\lambda_1 + \lambda_2 + \lambda_3 - 5\lambda_4 \ge 4$
$-\lambda_1 + 2\lambda_2 + \lambda_3 + \lambda_4 \ge 1$
$\lambda_i \ge 0;\quad i = 1,4$

$$\Rightarrow \lambda^* = \begin{Bmatrix} 7/3 \\ 5/3 \\ 0 \\ 0 \end{Bmatrix} \Rightarrow \boxed{L(\lambda^*) = 32} \Rightarrow$$

$$\Rightarrow \begin{cases} \nabla_x L(\mathbf{x}^*, \lambda^*) = \mathbf{c} + \mathbf{A}^T\lambda = 0 \Rightarrow \begin{Bmatrix} -4 \\ -1 \end{Bmatrix} + \begin{bmatrix} 1 & 1 & 1 & -5 \\ -1 & 2 & 1 & 1 \end{bmatrix} \begin{Bmatrix} 7/3 \\ 5/3 \\ 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \Rightarrow \mathbf{x}^* = \begin{Bmatrix} 4 \\ 2 \end{Bmatrix} \Rightarrow \boxed{f(\mathbf{x}^*) = 32} \\ x^*_1 - x^*_2 = 2; \qquad x^*_1 + 2x^*_2 = 8 \end{cases}$$

Quadratic Programming

A problem in Quadratic Programming is defined as:

Min $F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\boxed{\mathbf{A}}\mathbf{x} + \mathbf{bx}$
$\mathbf{x}$
Subject to $\mathbf{Bx} \le \mathbf{c}$
$\mathbf{x} \ge \mathbf{0}$

Hessian Matrix

Suppose n design variables and m constraints

The Lagrangian of the problem is equal to:

$$L(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^T\mathbf{Ax} + \mathbf{bx} + \lambda^T(\mathbf{Bx} - \mathbf{c}) \Rightarrow \nabla_x L(\mathbf{x}, \lambda) = \mathbf{Ax} + \mathbf{b} + \mathbf{B}^T\lambda = 0 \Rightarrow$$

$$\Rightarrow \mathbf{x} = -\mathbf{A}^{-1}(\mathbf{b} + \mathbf{B}^T\lambda) \Rightarrow L(\lambda) = -\frac{1}{2}\lambda^T\mathbf{D}\lambda - \mathbf{d}.\lambda \boxed{-\frac{1}{2}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}}$$

Constant term

where: $\mathbf{D} = \mathbf{BA}^{-1}\mathbf{B}^T$ e $\mathbf{d} = \mathbf{c} + \mathbf{BA}^{-1}\mathbf{b}$

Thus, the new problem can be solved using the duality concept:

Max $L(\lambda) = -\frac{1}{2}\lambda^T\mathbf{D}\lambda - \mathbf{d}.\lambda$
λ
$(\lambda \ge \mathbf{0})$

m design variables with simple constraints

SCILAB Implementation

37

SCILAB routine to solve quadratic programming: **QUAPRO**

[x,lagr,f]=quapro(Q,p,C,b,ci,cs,mi,x0) solves the problem:

Symmetric matrix (nxn)

$$\underset{\mathbf{x}}{\text{Min}} \quad f = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{p}^T\mathbf{x}$$

Subject to $\mathbf{C}_1\mathbf{x} = \mathbf{b}_1$ $(m_i \times n)$

$\mathbf{C}_2\mathbf{x} \leq \mathbf{b}_2$ $(m_d \times n)$

$\mathbf{c}_i \leq \mathbf{x} \leq \mathbf{c}_s$

$\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$ $(m_i + m_d) \times n$

$\mathbf{b} = [\mathbf{b}_1, \mathbf{b}_2]$ $1 \times (m_i + m_d)$

- x – vector of the optimal solution;
- lagr – vector of the Lagrange multipliers ($n+m_i+m_d$). If the value of a multiplier is zero, the corresponding constraint is not active; otherwise it is active at the optimal;
- f – optimal value of the function;
- m_i – number of equality constraints;
- x_0 – start guess of x; or "x_0=v" the calculation is started at the vertex of the domain; or "x_0=g" the initial value is arbitrary.

Prof. Dr. Emilio C. Nelli Silva

Constrained Nonlinear Optimization Algorithms

38

Classification:

- Direct Methods ➡ verify the constraints after they follow the unconstrained algorithm procedure.
  Ex.: Gradient Projection and Reduced Gradient Methods and Feasible Directions Method
- Indirect Methods ➡ transform the constrained problem into an unconstrained one.
  Ex.: Penalization Methods and Augmented Lagrange Method

Prof. Dr. Emilio C. Nelli Silva

Direct Methods

39

Gradients Projection and Reduced Gradient Methods

1) Initially we consider linear constraints:

$$\min_{\mathbf{x}} \quad f(\mathbf{x})$$

Subject to $g_j = \mathbf{a}_j^T \mathbf{x} - \mathbf{b}_j \geq 0$

$j = 1, \ldots, n_g$

Considering r active constraints:

$\mathbf{g}_a = \mathbf{N}^T\mathbf{x} - \mathbf{b} = 0$ where: $n_{ij} = a_{ij}$

Basic hypothesis: $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha \mathbf{s}$

Search on the boundary of the domain

$\mathbf{x}$ is located in the tangent subset relative to the active constraints, thus: $\mathbf{N}^T\mathbf{x}_{i+1} = \mathbf{N}^T\mathbf{x}_i + \alpha\mathbf{N}^T\mathbf{s} \Rightarrow \mathbf{b} = \mathbf{b} + \alpha\mathbf{N}^T\mathbf{s} \Rightarrow \mathbf{N}^T\mathbf{s} = 0$

The direction $\mathbf{s}$ is obtained after the solution of the problem:

$$\min_{\mathbf{s}} \quad \mathbf{s}^T \nabla f$$

Subject to $\mathbf{N}^T\mathbf{s} = 0$

$\mathbf{s}^T\mathbf{s} = 1$

$$L = \mathbf{s}^T\nabla f - \mathbf{s}^T\mathbf{N}\lambda - \mu\left(\mathbf{s}^T\mathbf{s} - 1\right) \Rightarrow \frac{\partial L}{\partial \mathbf{s}} = \nabla f - \mathbf{N}\lambda - 2\mu\mathbf{s} = 0 \Rightarrow$$

$$\Rightarrow \mathbf{N}^T\nabla f - \mathbf{N}^T\mathbf{N}\lambda = 0 \Rightarrow \lambda = \left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{N}^T\nabla f \Rightarrow$$

$$\Rightarrow s = \frac{1}{2\mu}\left[\mathbf{I} - \mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{N}^T\right]\nabla f = \frac{1}{2\mu}\mathbf{P}\nabla f$$

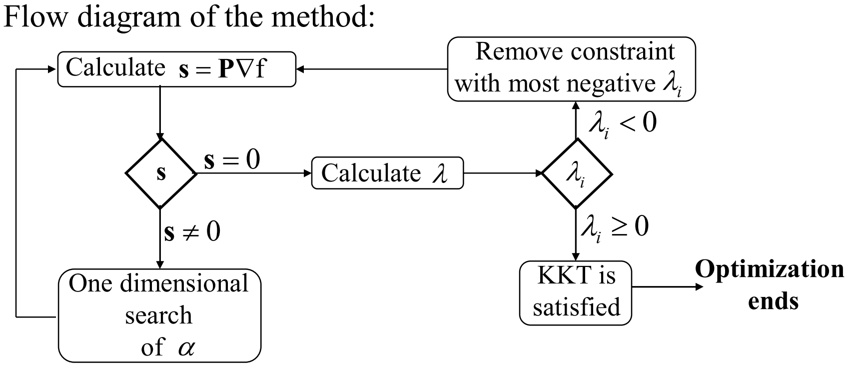
Where for arbitrary $\mathbf{w}$, $\mathbf{Pw}$ is the subset tangent to the active constraints, thus: $\mathbf{N}^T\mathbf{P}\mathbf{w} = \mathbf{N}^T[\mathbf{I} - \mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{N}^T]\mathbf{w} = 0$

Prof. Dr. Emilio C. Nelli Silva

Gradients Projection and Reduced Gradient Methods

40

Flow diagram of the method:



Verification of the violation of the constraints:

$$g_j = \mathbf{a}_j^T(\mathbf{x}_i + \alpha\mathbf{s}) - b_j \geq 0 \Rightarrow \alpha \leq -\frac{\left(\mathbf{a}_j^T\mathbf{x}_i - b_j\right)}{\mathbf{a}_j^T\mathbf{s}} = -\frac{g_j(\mathbf{x}_i)}{\mathbf{a}_j^T\mathbf{s}}$$

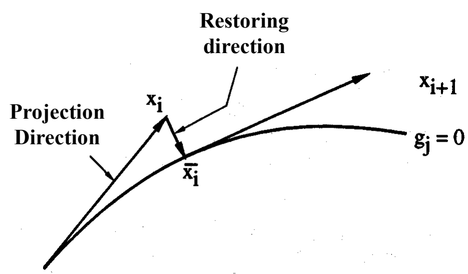
Valid if $\mathbf{a}_j^T\mathbf{s} < 0$, otherwise there is no superior limit of α due to the j-th constraint. The limit value of α is given by:

$$\bar{\alpha} = \min_{\substack{\alpha_j > 0 \\ \text{active constraints}}} \alpha_j$$

Prof. Dr. Emilio C. Nelli Silva

Gradients Projection and Reduced Gradient Methods

2) Considering nonlinear constraints
Linearize the constraint around x_i ➡ one-dimensional search crosses the boundary of the constraint ➡ correction



Correction:
$g_j \approx g_j(\mathbf{x}_i) + \nabla g_j^T(\bar{\mathbf{x}}_i - \mathbf{x}_i)$
Verify that:
$\bar{\mathbf{x}}_i - \mathbf{x}_i = -\mathbf{N}(\mathbf{N}^T\mathbf{N})^{-1}\mathbf{g}_a(\mathbf{x}_i)$

Active constraints

α is determined specifying a reduction in the γ objective function:

$$f(\mathbf{x}_i) - f(\mathbf{x}_{i+1}) \cong \nabla f^T(\mathbf{x}_i - \mathbf{x}_{i+1}) \approx \gamma f(\mathbf{x}_i) \Rightarrow \alpha^* = -\frac{\gamma f(\mathbf{x}_i)}{\mathbf{s}^T\nabla f} \Rightarrow$$

$$\Rightarrow \boxed{\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha^*\mathbf{s} - \mathbf{N}(\mathbf{N}^T\mathbf{N})^{-1}\mathbf{g}_a}$$

Example

$$\min_{x_1,x_2} \quad f = 3x_1 + \sqrt{3}x_2$$

Subject to
$$g_1 = 3 - \frac{18}{x_1} - \frac{6\sqrt{3}}{x_2} \geq 0$$
$$g_2 = x_1 - 5.73 \geq 0$$
$$g_3 = x_2 - 7.17 \geq 0$$

$$\mathbf{x}_0 = (11{,}61;7{,}17) \Rightarrow g_1 = 0,\ g_3 = 0 \Rightarrow f = 47{,}25 \Rightarrow$$

$$\Rightarrow \nabla f = \begin{Bmatrix} 3 \\ \sqrt{3} \end{Bmatrix};\ \nabla g_1 = \begin{Bmatrix} 0{,}1335 \\ 0{,}2021 \end{Bmatrix};\ \nabla g_3 = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix};\ N = \begin{bmatrix} 0{,}1335 & 0 \\ 0{,}2021 & 1 \end{bmatrix} \Rightarrow$$

$$\Rightarrow \mathbf{P} = 0 \Rightarrow \mathbf{s} = 0 \Rightarrow \lambda = \begin{Bmatrix} 22{,}47 \\ -2{,}798 \end{Bmatrix} \Rightarrow$$ eliminate g_3, thus:

$$N = \begin{bmatrix} 0{,}1335 \\ 0{,}2021 \end{bmatrix} \Rightarrow \mathbf{P} = \begin{bmatrix} 0{,}6962 & -0{,}46 \\ -0{,}46 & 0{,}3036 \end{bmatrix} \Rightarrow \mathbf{s} = -\mathbf{P}\nabla f = \begin{Bmatrix} -1{,}29 \\ 0{,}854 \end{Bmatrix}$$

For a 5% reduction in the objective function (γ=0,05): $\alpha^* = \dfrac{0{,}05 \times 47{,}25}{\begin{bmatrix} -1{,}29 & 0{,}854 \end{bmatrix}\begin{Bmatrix} 3 \\ \sqrt{3} \end{Bmatrix}} = 0{,}988$

Since there if no violation in the constraint ➡ there is no need for correction:

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha^*\mathbf{s} = \begin{Bmatrix} 11{,}61 \\ 7{,}17 \end{Bmatrix} + 0{,}988\begin{Bmatrix} -1{,}29 \\ 0{,}854 \end{Bmatrix} = \begin{Bmatrix} 10{,}34 \\ 8{,}01 \end{Bmatrix} \Rightarrow f(\mathbf{x}_1) = 44{,}89;$$

$g(\mathbf{x}_1) = -0{,}0382 \Rightarrow g_1$ constraint active ➡ $\mathbf{N} = \nabla g_1 = \begin{Bmatrix} 0{,}1684 \\ 0{,}1620 \end{Bmatrix} \Rightarrow$

$$\Rightarrow \mathbf{P} = \begin{bmatrix} 0{,}4806 & -0{,}4996 \\ -0{,}4996 & 0{,}5194 \end{bmatrix} \Rightarrow$$

Example

$$\Rightarrow \mathbf{s} = -\mathbf{P}\nabla f = \begin{Bmatrix} -0,5764 \\ 0,5991 \end{Bmatrix}; \ \gamma = 0,025 \Rightarrow \alpha = -\frac{0,025 \times 44,89}{\begin{bmatrix} -0,5764 & 0,599 \end{bmatrix}\begin{Bmatrix} 3 \\ \sqrt{3} \end{Bmatrix}} = 1,62$$

Since g_1 is violated ➡ there is correction

$$\mathbf{g}_a = -0,0382 \Rightarrow -\mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{g}_a = \begin{Bmatrix} 0,118 \\ 0,113 \end{Bmatrix} \Rightarrow \mathbf{x}_2 = \mathbf{x}_1 + \alpha^*\mathbf{s} - \mathbf{N}\left(\mathbf{N}^T\mathbf{N}\right)^{-1}\mathbf{g}_a = \begin{Bmatrix} 10,34 \\ 8,01 \end{Bmatrix} +$$

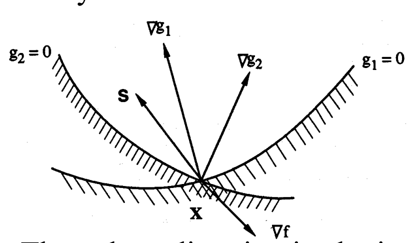
$$-1,62\begin{Bmatrix} 0,576 \\ -0,599 \end{Bmatrix} + \begin{Bmatrix} 0,118 \\ 0,113 \end{Bmatrix} = \begin{Bmatrix} 9,52 \\ 9,10 \end{Bmatrix} \Rightarrow f(\mathbf{x}_2) = 44,32; \ g_1(\mathbf{x}_2) = -0,0328$$

$$\mathbf{x}^* = \begin{Bmatrix} 9,464 \\ 9,464 \end{Bmatrix} \Rightarrow f(\mathbf{x}^*) = 44,78 \qquad \text{(Optimal solution)}$$

Direct Methods

The Method of Feasible Directions

Aims to stride away from the boundary of the constraints. Start with the boundary of the feasible domain.



Active constraints

Viable direction: $\mathbf{s}^T\boxed{\nabla g_j} > 0$

Useful direction: $\mathbf{s}^T\nabla f < 0$

Thus, the **s**-direction is obtained through the solution of the optimization problem:

Linear Programming

$$\begin{array}{ll} \underset{\mathbf{s}}{\text{Min}} & \beta \\ \text{Subject to} & -\mathbf{s}^T\nabla g_j + \theta_j\beta \leq 0 \\ & \mathbf{s}^T\nabla f + \beta \leq 0 \qquad \theta_j \geq 0 \\ & |s_i| \leq 1 \end{array}$$

If $\beta_{max} = 0 \Rightarrow$ KKT conditions are satisfied.

Example

$$\min_{x_1, x_2} \quad f = 3x_1 + \sqrt{3}x_2$$

Subject to
$$g_1 = 3 - \frac{18}{x_1} - \frac{6\sqrt{3}}{x_2} \geq 0$$
$$g_2 = x_1 - 5.73 \geq 0$$
$$g_3 = x_2 - 7.17 \geq 0$$

$$\mathbf{x}_0^T = (11{,}61;7{,}17) \Rightarrow g_1 = 0,\ g_3 = 0 \Rightarrow f = 47{,}25 \Rightarrow$$

$$\Rightarrow \nabla f = \begin{Bmatrix} 3 \\ \sqrt{3} \end{Bmatrix};\ \nabla g_1 = \begin{Bmatrix} 0{,}1335 \\ 0{,}2021 \end{Bmatrix};\ \nabla g_3 = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$$

Selecting $\theta_1 = \theta_2 = 1$, it is found:

$$\min_{s_1, s_2} \quad \beta$$

Subject to
$$-0{,}1335 s_1 - 0{,}2021 s_2 + \beta \leq 0$$
$$-s_2 + \beta \leq 0$$
$$3s_1 + \sqrt{3}s_2 + \beta \leq 0$$
$$-1 \leq s_1 \leq 1$$
$$-1 \leq s_2 \leq 1$$

Solution: $s_1 = -0{,}6172;\ s_2 = 1 \Rightarrow$

One-dimensional search:

$$\Rightarrow \mathbf{x}_1 = \begin{Bmatrix} 11{,}61 \\ 7{,}17 \end{Bmatrix} + \alpha \begin{Bmatrix} -0{,}6172 \\ 1 \end{Bmatrix}$$

α is limited by $g_1 \Rightarrow \alpha < 5{,}385 \Rightarrow$

$$\Rightarrow \alpha = 5{,}385 \Rightarrow \mathbf{x}_1 = \begin{Bmatrix} 8{,}29 \\ 12{,}56 \end{Bmatrix} \Rightarrow f(\mathbf{x}_1) = 46{,}62$$

Next iteration (only g_1 active):

$$\nabla g_1 = \begin{Bmatrix} -\dfrac{18}{x_1^2} \\ -\dfrac{6\sqrt{3}}{x_2^2} \end{Bmatrix} = \begin{Bmatrix} 0{,}2619 \\ 0{,}0659 \end{Bmatrix};$$

$$\nabla f = \begin{Bmatrix} 3 \\ 3\sqrt{3} \end{Bmatrix}$$

Example

$$\min_{s_1, s_2} \quad \beta$$

Subject to
$$-0{,}2619 s_1 - 0{,}0659 s_2 + \beta \leq 0$$
$$3s_1 + \sqrt{3}s_2 + \beta \leq 0$$
$$-1 \leq s_1 \leq 1$$
$$-1 \leq s_2 \leq 1$$

Solution: $s_1 = 0{,}5512;\ s_2 = -1 \Rightarrow$

One dimensional search:

$$\Rightarrow \mathbf{x}_2 = \begin{Bmatrix} 8{,}29 \\ 12{,}56 \end{Bmatrix} + \alpha \begin{Bmatrix} 0{,}5512 \\ -1 \end{Bmatrix}$$

α is limited by $g_1 \Rightarrow \alpha < 4{,}957 \Rightarrow$

$$\Rightarrow \alpha = 4{,}957 \Rightarrow \mathbf{x}_2 = \begin{Bmatrix} 11{,}02 \\ 7{,}60 \end{Bmatrix} \Rightarrow f(\mathbf{x}_2) = 46{,}22$$

$$\mathbf{x}^* = \begin{Bmatrix} 9{,}464 \\ 9{,}464 \end{Bmatrix} \Rightarrow f(\mathbf{x}^*) = 44{,}78$$

(Optimal Solution)