Differentiable manifolds

O.M. Kiselev
o.kiselev@innopolis.ru

Innopolis university

March 10, 2023

Least squares approximation

Let us consider the problem to find the best approximation for the set of points $\{(x_i, y_i)\}_{i=1}^{n}$ by the straight line $y = a_1 x + a_2$ the problem to find optimal values of (a_1, a_2).
Define the sum of squared residuals:

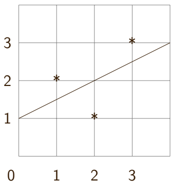
$$S(k, b) = \sum_{i=1}^{n} (y_i - a_1 x_i - a_2)^2.$$

The minimum of the function with respect to parameters k, b defines the best approximation. The conditions are:

$$\frac{\partial S}{\partial a_1} = 2 \sum_{i=1}^{n} x_i (y_i - a_1 x_i - a_2) = 0,$$

$$\frac{\partial S}{\partial a_2} = 2 \sum_{i=1}^{n} (y_i - a_1 x_i - a_2) = 0.$$

Least squares approximation. Example



Consider
the set of points $(1,2),(2,1),(3,3)$.
Define the sum of squared residuals:

$$S(a_1,a_2)=(a_1+a_2-2)^2+(2a_1+a_2-1)^2+(3a_1+a_2-3)^2.$$

$$\frac{\partial S}{\partial a_1}=28a_1+12a_2-26=0,\ \ \frac{\partial S}{\partial a_2}=12a_1+6a_2-12=0.$$

$$28a_1+12a_2=26,\quad 12a_1+6a_2=12.$$

$$a_1=1/2,\ \ a_2=1\Rightarrow y=\frac{1}{2}x+1.$$

Image correction

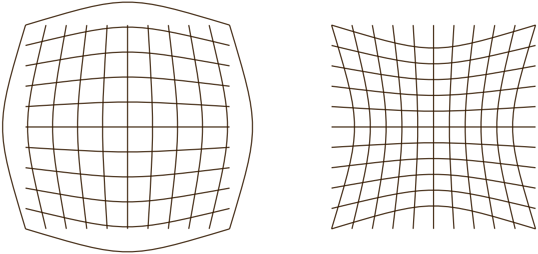


Figure: Image distortions: barrel and pincushion

The task is to find maps $X = X(x, y),\ \ Y = Y(x, y)$ which minimize the sum of residuals:

$$S = \sum_{i=0}^{N} \sum_{j=0}^{M} (X(x_{ij}, y_{ij}) - i)^2 + (Y(x_{ij}, y_{ij}) - j)^2 \to \min . \quad (1)$$

Linear image correction

Here we use the linear mapping to correct the image

$$X(x, y) = ax + by + c,$$
$$Y(x, y) = ux + vy + w.$$

Problem is find coefficients a, b, c, u, v, w to minimize the sum of squared residuals.

Linear image correction

The necessary conditions for the coefficients:

$$
\begin{aligned}
2a\sum_{i,j} x_{ij}^2 + 2b\sum_{i,j} x_{ij}y_{ij} + 2c\sum_{i,j} x_{ij} - 2\sum_{i,j} ix_{ij} &= 0,\\
2a\sum_{i,j} x_{ij}y_{ij} + 2b\sum_{i,j} y_{ij}^2 + 2c\sum_{i,j} y_{ij} - 2\sum_{i,j} iy_{ij} &= 0,\\
2a\sum_{i,j} x_{ij} + 2b\sum_{i,j} y_{ij} + 2cMN - M(N-1)N &= 0,\\
2u\sum_{i,j} x_{ij}^2 + 2v\sum_{i,j} x_{ij}y_{ij} + 2w\sum_{i,j} x_{ij} - 2\sum_{i,j} jx_{ij} &= 0,\\
2u\sum_{i,j} x_{ij}y_{ij} + 2v\sum_{i,j} y_{ij}^2 + 2w\sum_{i,j} y_{ij} - 2\sum_{i,j} jy_{ij} &= 0,\\
2u\sum_{i,j} x_{ij} + 2v\sum_{i,j} y_{ij} + 2wMN - NM(M-1) &= 0.
\end{aligned}
$$

Linear image correction

Let change the variables:

$$\tilde{X}_3 = \sum_{i,j} x_{ij}^2, \quad \tilde{X}_2 = \sum_{i,j} i x_{ij}, \quad \tilde{X}_1 = \sum_{i,j} j x_{ij}, \quad \tilde{X}_0 = \sum_{i,j} x_{ij}$$

$$\tilde{Z} = \sum_{i,j} x_{ij} y_{ij}, \quad \tilde{Y}_3 = \sum_{i,j} y_{ij}^2, \quad \tilde{Y}_2 = \sum_{i,j} i y_{ij}, \quad \tilde{Y}_1 = \sum_{i,j} j y_{ij}, \quad \tilde{Y}_0 = \sum_{i,j} y_{ij}.$$

Here the letters with tildes are the sums of known expressions of x_{ij} and y_{ij}.

Linear image correction

Then we can rewrte the equations for a, b, c

$$
\begin{aligned}
2a\tilde{X}_3 + 2b\tilde{Z} + 2c\tilde{X} - 2\tilde{X}_2 &= 0,\\
2a\tilde{Z} + 2b\tilde{Y}_3 + 2c\tilde{Y} - 2K\tilde{Y}_2 &= 0,\\
2a\tilde{X} + 2b\tilde{Y} + 2cMN - MN(N-1) &= 0
\end{aligned}
$$

and for u, v, w.

$$
\begin{aligned}
2u\tilde{X}_3 + 2v\tilde{Z} + 2w\tilde{X} - 2\tilde{X}_1 &= 0,\\
2u\tilde{Z} + 2v\tilde{Y}_2 + 2w\tilde{Y} - 2\tilde{Y}_1 &= 0,\\
2u\tilde{X} + 2v\tilde{Y} + 2wMN - NM(M-1) &= 0.
\end{aligned}
$$

These solutions define the linear maps $X(x, y)$ and $Y(x, y)$.

Linear image correction $S \sim 8.5 \rightarrow S = 4.23$

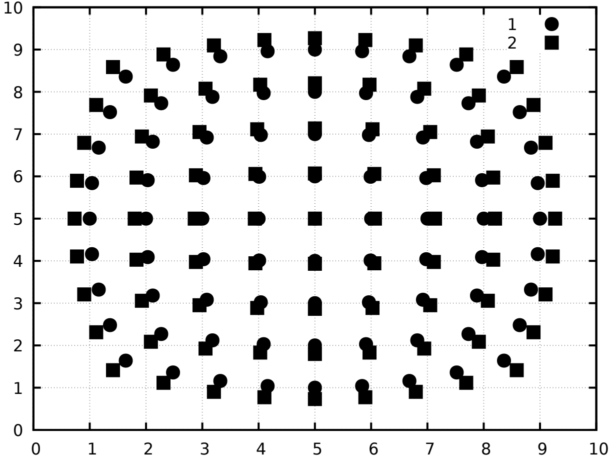


Figure: Linear correction of barrel distorsion.

Cubic image correction

Consider a cubic map:

$$
\begin{aligned}
X(x,y) &= a_0 + a_1 x + a_2 y + b_{20}x^2 + b_{11}xy + b_{02}y^2 + \\
&\qquad c_{30}x^3 + c_{21}x^2y + c_{12}xy^2 + c_{03}y^3, \\
Y(x,y) &= u_0 + u_1 x + u_2 y + v_{20}x^2 + v_{11}xy + v_{20}y^2 + \\
&\qquad w_{30}x^3 + w_{21}x^2y + w_{12}xy^2 + w_{03}y^3.
\end{aligned}
$$

The necessary conditions are

$$
\begin{aligned}
&\frac{\partial S}{\partial a_i} = 0, \quad i = 0,1,2; \quad \frac{\partial S}{\partial u_i} = 0, \quad i = 0,1,2; \\
&\frac{\partial S}{\partial b_{ij}} = 0 \quad 0 \geq i,j, \quad i+j = 2; \quad \frac{\partial S}{\partial v_{ij}} = 0 \quad 0 \geq i,j, \quad i+j = 2; \\
&\frac{\partial S}{\partial c_{ij}} = 0 \quad 0 \geq i,j, \quad i+j = 3; \quad \frac{\partial S}{\partial b_{ij}} = 0 \quad 0 \geq i,j, \quad i+j = 3.
\end{aligned}
$$

Cubic image correction $S \sim 8.5 \rightarrow S \sim 0.08$

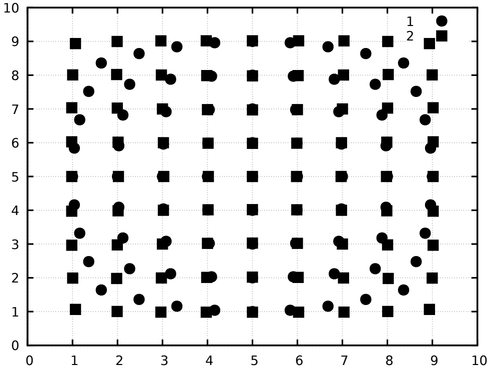


Figure: Cubic correction of barrel distorsion.

The cubic map leads to the sum of residuals from $S \sim 8.5$ to $S \sim 0.08$ which better wthat the linear correction: $S = 4.23$.

A gradient-wise descent

Seeking of a local minima looks like a descent to the lowest point in a neighborhood. The best direction to the descent opposites to the gradient.

Example

Define the gradient of the function

$$f(\vec{x}) = x_1^2 - 3x_1x_2 + x_2^4,$$

in the point $(1, 2)$.
The partial derivatives are:

$$\frac{\partial f}{\partial x_1} = 2x_1 - 3x_2, \quad \frac{\partial f}{\partial x_2} = -3x_1 + 4x_2^3.$$

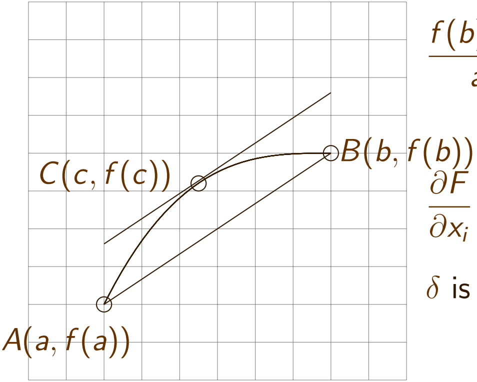
$$\vec{\nabla} f|_{(1,2)} = (-4, 29).$$

Typical steps of the gradient descent

Consider an algorithm of the gradient descent. Let Δ be a length of the step in opposite of the gradient.

- Let a current point $x_1, \ldots, x_n$.
- Calculate the gradient F at the current point.
- A step in direction opposite to the gradient $\vec{X} = \vec{x} - \Delta \mathrm{grad}(F)$.
- Check $F(\vec{X}) < F(\vec{x})$.
- If the condition fulfills then new position $\vec{x} = \vec{X}$.
- Another case a local minimum in the distance less than Δ.

Calculation of the partial derivatives using the Lagrange theorem



$$\frac{f(b) - f(a)}{a - b} = f'(c), \quad c \in (a, b),$$

$$\frac{\partial F}{\partial x_i} \sim \frac{F(x_i + \delta) - F(x_i - \delta)}{2\delta}.$$

δ is small.

The fastest gradient-wise descent

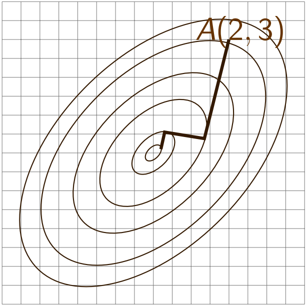
Change the gradient descent method for the opposite gradient direction as function of the variable Δ:

$$\Phi(\Delta) = F(\vec{x} - \vec{\nabla} F(\vec{x}) \cdot \Delta), \quad \{x_1, \ldots, x_N\} = \text{const}.$$

So we seek the minimum of the one dimension function $\Phi(\Delta)$ on the given direction.

1. Define an interval $\Delta \in [0, b]$ such that $\Phi(\Delta) \leq \Phi(0)$.
2. Find a minimum $\Phi(\Delta^*)$ on $\Delta \in [0, b]$ using for example a bisection method.
3. The point $\vec{X} = \vec{x} - \vec{\nabla} F(\vec{x}) \cdot \Delta^*$ is considered as next position for the next step.
4. If $||\vec{X} - \vec{x}|| > \delta$, then this process repeats.

An example



Consider the fastest gradient descent for the function

$$f(x_1, x_2) = (x_1+x_2)^2+3(x_1-x_2)^2.$$

The level curves are ellipses with big semi axis along the straight line $x_1 = x_2$ and the minimum is $(0, 0)$.

An example

Let the initial point is $(2,3)$ then $f(2,3)=28$. Find the derivatives and the gradient:

$$\frac{\partial f}{\partial x_1}=8x_1-4x_2,\ \frac{\partial f}{\partial x_2}=-4x_1+8x_2,\ \vec{\nabla} f|_{(2,3)}=(4,16);$$

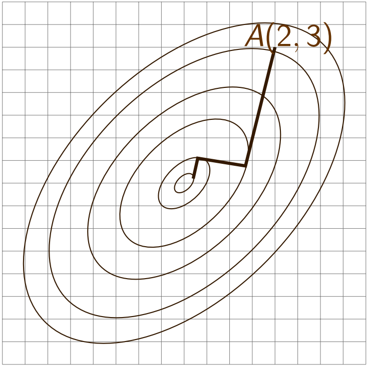
Then the descent direction starting the point $(2,3)$:
$x_1=2-4t, \quad x_2=3-16t, \quad t>0$. The minimized one-dimensional function is:

$$f(x_1(t),x_2(t))=832t^2-272t+28=832\left(t-\frac{272}{2\cdot 832}\right)^2+\frac{75}{13}.$$

$$t_1=\frac{272}{2\cdot 832}\sim 0.16346,\ x_1=2-4t_1\sim 2-4\cdot 0.16346\sim 1.35,$$

$$x_2=3-16t_1\sim 3-16\cdot 0.16346\sim 0.38,\ f=\frac{75}{13}\sim 5.77.$$

An example



The track of
minimizing for the function
$f = (x_1 + x_2)^2 + 3(x_1 - x_2)^2$
is shown of the picture.
All
turning points lie on tangent
line for the level curves.

A global minimum

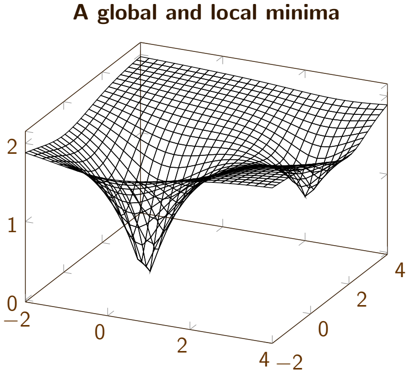


Figure: The example of the surface with two minima which are a global and local ones. If one starts the gradient descent near the local minimum, then one does not reach the global minimum.

Primary concepts for definition of a differentiable manifold

- $a_1x_1 + a_2x_2 + c = 0$, is a one function of one variable: $x_1(x_2) = -\frac{1}{a_1}(a_2x_2 + c)$ or $x_2(x_1) = -\frac{1}{a_2}(a_1x_1 + c)$. Both forms are appropriated if $a_{1,2} \neq 0$.
- $a_1x_1 + a_2x_2 + a_3x_3 + c = 0$ is a a function of two variables: $x_k = -\frac{1}{a_k}\left(\sum_{n\neq k} a_nx_n + c\right)$.
- The following two equals define one dimensional function.

$$a_1x_1 + a_2x_2 + a_3x_3 + c = 0,\ \ b_1x_1 + b_2x_2 + b_3x_3 + d = 0.$$

- In a general case the m equalities of N variables define $N - m$ dimensional implicit function:

$$f_k(x_1, \ldots, x_N) = 0,\ k = 1, \ldots, m.$$

Primary concepts for definition of differentiable manifolds

- Does the formula $\sum_{k=1}^{3} x_k^2 - R^2 = 0$ define a two-dimensional function?

$$x_i = +\sqrt{R^2 - \sum_{k=1,k\neq i}^{3} x_k^2}, \; x_i = -\sqrt{R^2 - \sum_{k=1,k\neq i}^{3} x_k^2}.$$

The answer looks like NO! because one obtains two different values for x_i for the same set of coordinates $\{x_k\}_{i\neq k}$.

Primary concepts for definition of differentiable manifolds

- Another point of view for the spherical coordinates:

  $$x_1 = R\sin(\theta)\cos(\phi),\ x_2 = R\sin(\theta)\sin(\phi),\ x_3 = R\cos(\theta).$$

  So we can see that in the spherical coordinate system one obtain one-to-one map $[0,\pi]\times[0,2\pi) \to$ a set of x_1, x_2, x_3.

We need a generalization for the function definition.

An example

Consider

$$x^2 + y^2 + z^2 - 1 = 0,$$

A normal vector $\vec{N}$ at a point $A = (x_0, y_0, z_0)$ looks like $\vec{N} = (2x_0, 2y_0, 2z_0)$.
The tangent plain at the point A is follows:$2x_0(x - x_0) + 2y_0(y - y_0) + 2z_0(z - z_0) = 0$.
Without loss of a generality let's assume $x_0 \neq 0$. Define $\vec{e}_1$:

$$(\vec{N}, \vec{\eta}) = 0,\ \vec{\eta} = (-z_0, 0, x_0),$$
$$\vec{\zeta}: \ \vec{\zeta} = \vec{\eta} \times \vec{N} = (-x_0 y_0, x_0^2 + z_0^2, -y_0 z_0),$$
$$(\vec{\eta}, \vec{\zeta}) = z_0 x_0 y_0 - x_0 y_0 z_0 = 0.$$

The $\vec{\eta}$ and $\vec{\zeta}$ define an orthogonal basis in the tangent plain.

An example

New coordinates (ξ, η, ζ):

$$\begin{pmatrix} 2x_0 & 2y_0 & 2z_0 \\ -z_0 & 0 & x_0 \\ -x_0 y_0 & x_0^2 + z_0^2 & -y_0 z_0 \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} = \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix}$$

A determinant of the matrix $y_0^2 - 1$.
The inverse transform:

$$\begin{pmatrix} x_0 & \frac{z_0}{y_0^2 - 1} & \frac{x_0 y_0}{y_0^2 - 1} \\ y_0 & 0 & 1 \\ z_0 & \frac{x_0}{1 - y_0^2} & \frac{y_0 z_0}{y_0^2 - 1} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix}$$

An example

For any point A of the implicit function

$$x^2 + y^2 + z^2 - 1 = 0$$

one can define a new coordinate system and rewrite the implicit function in the explicit one in new coordinates.

Theorem about explicit form of the function

Let's consider

$$f_k(x_1, \ldots, x_n) = 0, k \in \{1, \ldots, M\},\ n \in \{1, \ldots, N\},$$

where all f_k are continuously differentiable functions at the origin.
If a rank of the matrix

$$S = \left(\frac{\partial f_k}{\partial x_n}\right),\ s_{k,n} = \frac{\partial f_k}{\partial x_n}$$

is equal M at the origin then exists a neighborhood ofthe origin, where the implicit function can be rewritten in an explicit form as a function of $N - M$ independent variables.

Sketch of proof

Let's change the variables:
$\xi_k = f(x_1, \ldots .x_N)$, $k \in \{1, \ldots, M\}$, Consider the matrix $M \times M$

$$\tilde{S} = \left(\frac{\partial f_k}{\partial x_n} \right), \quad \mathrm{Rank}(\tilde{S}) = M.$$

Then $\exists \epsilon >,\ x_k = \phi_k(\vec{\xi}, x_{M+1}, \ldots, N)$, $\forall \vec{\xi} : ||\vec{\xi}|| < \epsilon$.

$$d\xi_k \equiv \sum_{n=1}^{M} \frac{\partial f_k}{\partial x_n} dx_n,$$

Sketch of proof

Then one can rewrite

$$x_k = \sum_{n=1}^{M} \frac{\partial \phi_k}{\partial \xi_n} \xi_k + o(||\xi||).$$

Hence one can rewrite:

$$x_k = F_k(x_{M+1}, \ldots, x_N), k \in \{1, 2, \ldots, M\}.$$

A definition of differentiable manifold

The set in N dimensional space is called $N-M$ dimensional differentiable manifold if for any point A of the set exists neighborhood of the A such that $\exists \epsilon > 0$ and the manifold can be defined by

$$x_k = F_k(x_1, \ldots, x_{N-M}) = 0,\ k \in \{M, \ldots, N\},\ \forall x : ||x|| < \epsilon.$$

The set of maps covered all range of the variables is called an atlas.

Definition for a Jacobian

Let's consider the changing of variables:

$$y_k = f_k(x).$$

the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ & \dots & \\ \frac{\partial f_k}{\partial x_1} & \dots & \frac{\partial f_k}{\partial x_n} \end{pmatrix}$$

is called Jacobian.

Example

$$y_1 = x_1^4 + x_1^2 x_2 + x_3, y_2 = x_2,\ \ y_3 = x_3.$$

$$\frac{\partial(y_1, y_2, y_3)}{\partial(x_1, x_2, x_3)} = \begin{pmatrix} 4x_1^3 + 2x_1x_2 & x_1^2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Summary

Least squares approximation

A gradient-wise descent

The fastest gradient-wise descent

Differentiable manifolds