CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Computing@CIT

Decision Analytics

Lecture 4: Least-squares estimation with implicit constraints (Gauss-Helmert-Model)

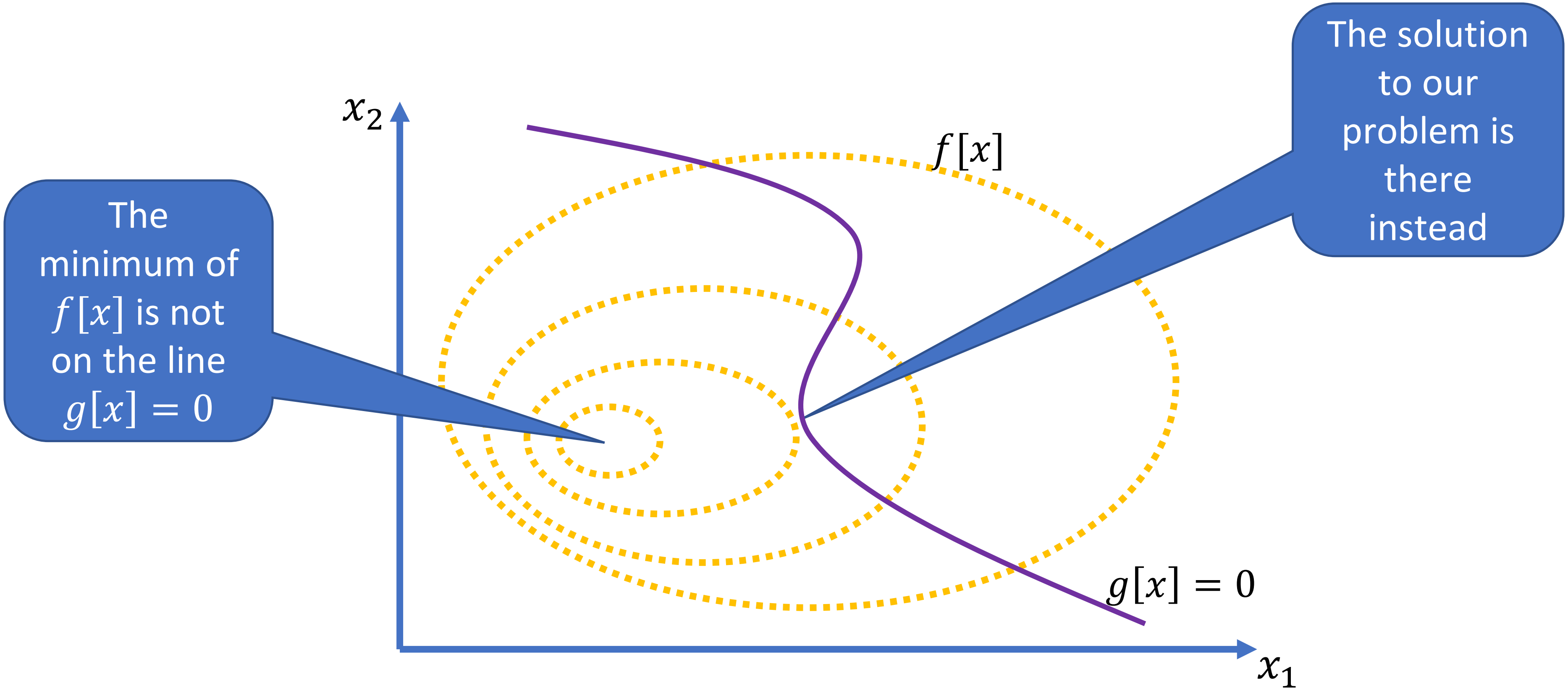
Lagrange Parameters

- We want to find the maximum of a function

$$f[x]$$

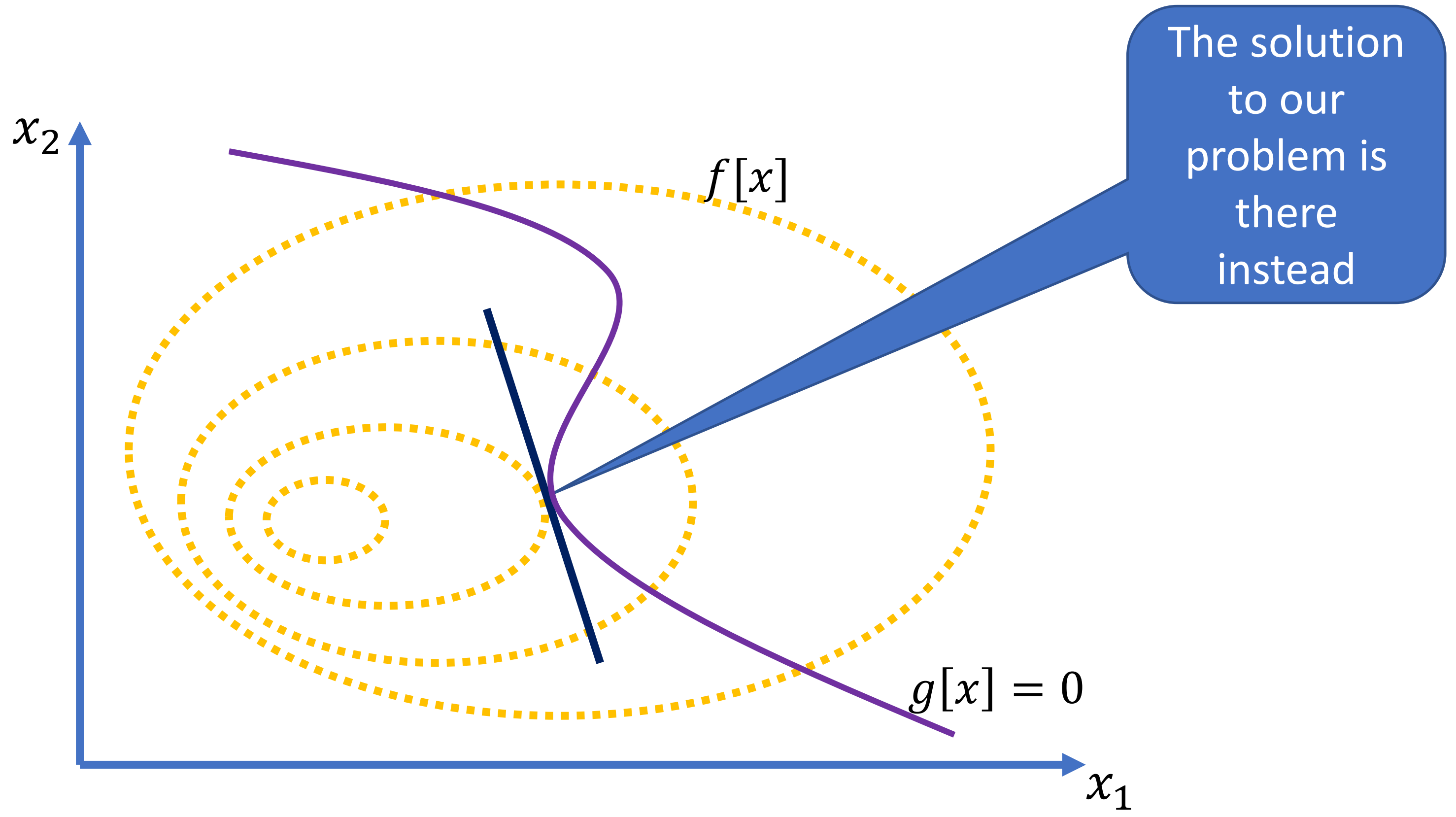
subject to the implicit constraints

$$g[x] = 0$$



Lagrange Parameters

- A necessary condition for the solution is that the tangent to the level-set of the objective function f is the same as the tangent to the constraint function g



Lagrange Parameters

- A necessary condition for the solution is that the tangent to the level-set of the objective function f is the same as the tangent to the constraint function g
- In other words, we require that the gradients of f and g point in the same direction

$$\nabla f = \lambda \nabla g$$

- Note, that the magnitude of the gradient of f and g can be different, so we need a factor λ (called the Lagrange multiplier)
- This argument can be extended towards multiple constraints, in which case we need as many Lagrange parameters as constraints

$$\nabla f = \sum_i \lambda_i \nabla g_i$$

Lagrange Parameters

- In summary, finding the maximum of a function

$$f[x]$$

subject to the implicit constraints

$$g[x] = 0$$

- Can be restated as finding the solution to the equation system

$$\nabla f - \sum_i \lambda_i \nabla g_i = 0$$

- Or equivalent as finding a stationary point to the Lagrange function

$$L[x, \lambda] = f[x] - \lambda^T g[x]$$

Gauss-Helmert-Model

- Given a vector of observations $\bar{l} = \left(\bar{l}_1, \dots, \bar{l}_N\right)^T$
- And a model function $g[p, l] = 0$
- We want to find the parameters $p = (p_1, \dots, p_U)^T$
- Subject to the constraints $h[p] = 0$
- So that the difference between the "true" values l (that do not violate the model constraints) and the actual observations $\bar{l}$ is minimised

- This problem can be stated as minimising the Mahalanobis distance

$$\Omega = \left(\overline{l - l}\right)^T C^{-1}\left(l - \bar{l}\right)$$

- Subject to the constraints

$$g[p, l] = 0$$

and

$$h[p] = 0$$

Gauss-Helmert-Model

- Approximating the constraints with the local Taylor expansion at the approximate values l_0 and p_0 we get the linearised constraints

$$g[p,l] \approx g[p_0,l_0] + A(\mathrm{p}-\mathrm{p}_0) + B(\mathrm{l}-\mathrm{l}_0) = \mathrm{g}_0 + \mathrm{A}\Delta p + B\Delta l = 0$$

and

$$h[p] \approx h[p_0] + H(p-p_0) = h_0 + H\Delta p = 0$$

- Substituting $\Delta l = l - l_0$ into the objective function we get

$$\Omega = \left(l-\bar{l}\right)^T C^{-1}\left(l-\bar{l}\right) = \left(l_0-\bar{l}+\Delta l\right)^T C^{-1}\left(l_0-\bar{l}+\Delta l\right)$$

- The Lagrange function is now

$$L[\Delta p, \Delta l, \lambda, \mu] = \Omega[\Delta l] + 2\lambda^T(g_0 + A\Delta p + B\Delta l) + 2\mu^T(h_0 + H\Delta p)$$

Gauss-Helmert-Model

- To find the solution we calculate the partial derivatives of the Lagrange function

$$L = \left(l_0 - \bar{l} + \Delta l\right)^T C^{-1}\left(l_0 - \bar{l} + \Delta l\right) + 2\lambda^T(g_0 + A\Delta p + B\Delta l) + 2\mu^T(h_0 + H\Delta p)$$

- Which are

$$\frac{\partial}{\partial \Delta p} L[\Delta p, \Delta l, \lambda, \mu] = 2A^T\lambda + 2H^T\mu$$

$$\frac{\partial}{\partial \Delta l} L[\Delta p, \Delta l, \lambda, \mu] = 2C^{-1}\left(l_0 - \bar{l}\right) + 2C^{-1}\Delta l + 2B^T\lambda$$

$$\frac{\partial}{\partial \lambda} L[\Delta p, \Delta l, \lambda, \mu] = 2g_0 + 2A\Delta p + 2B\Delta l$$

$$\frac{\partial}{\partial \mu} L[\Delta p, \Delta l, \lambda, \mu] = 2h_0 + 2H\Delta p$$

Gauss-Helmert-Model

- Setting all these partial derivatives to zero and dividing out the common factor 2 we obtain 4 linear equations

$$A^T\lambda + H^T\mu = 0$$
$$C^{-1}\left(l_0 - \bar{l}\right) + C^{-1}\Delta l + B^T\lambda = 0$$
$$g_0 + A\Delta p + B\Delta l = 0$$
$$h_0 + H\Delta p = 0$$

- Multiplying the second equation with C we get

$$\Delta l = \bar{l} - l_0 - CB^T\lambda$$

- Substituting this into the third equation yields

$$BCB^T\lambda = g_0 + A\Delta p + B\left(\bar{l} - l_0\right)$$

- Which we can solve for λ to get

$$\lambda = (BCB^T)^{-1}\left(g_0 + A\Delta p + B\left(\bar{l} - l_0\right)\right)$$

Gauss-Helmert-Model

- Finally we substitute λ into the first equation and get

$$A^T(BCB^T)^{-1}\left(g_0 + A\Delta p + B\left(\bar{l} - l_0\right)\right) + H^T\mu = 0$$

- Which together with the fourth equation

$$h_0 + H\Delta p = 0$$

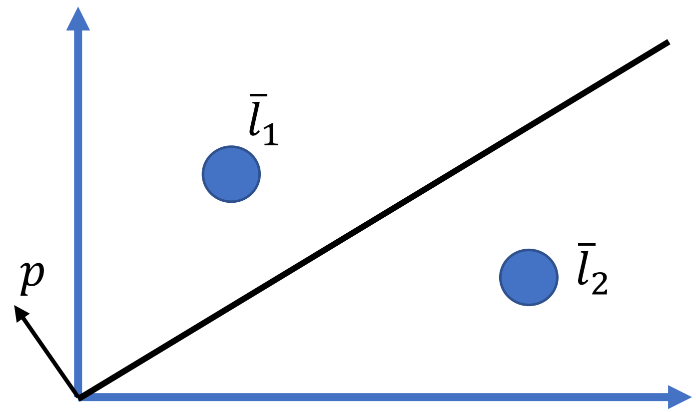
- Can be re-arranged into the Normal equation system

$$\begin{pmatrix} A^T(BCB^T)^{-1}A^T & H^T \\ H & 0 \end{pmatrix}\begin{pmatrix} \Delta p \\ \mu \end{pmatrix} = \begin{pmatrix} A^T(BCB^T)^{-1}\left(B\left(l_0 - \bar{l}\right) - g_0\right) \\ -h_0 \end{pmatrix}$$

Gauss-Helmert-Model

- The final iterative algorithm is as follows
  1. Choose good initial parameters p_0 and initialise $l_0 = \bar{l}$
  2. Calculate the Taylor expansion of g and h at p_0 and l_0 to obtain the Jacobian matrices A, B, and H
  3. Solve the Normal equation system to obtain Δp:
  $$\begin{pmatrix} \Delta p \\ \mu \end{pmatrix} = \begin{pmatrix} A^T (BCB^T)^{-1} A^T & H^T \\ H & 0 \end{pmatrix}^{-1} \begin{pmatrix} A^T (BCB^T)^{-1} \left(B\left(l_0 - \bar{l}\right) - g_0 \right) \\ -h_0 \end{pmatrix}$$
  4. Calculate
  $$\lambda = (BCB^T)^{-1} \left(g_0 + A\Delta p + B\left(\bar{l} - l_0\right)\right)$$
  5. Calculate
  $$\Delta l = \bar{l} - l_0 - CB^T \lambda$$
  6. Calculate the parameter estimate and adjusted observation
  $$\hat{p} = p_0 + \Delta p$$
  $$\hat{l} = l_0 + \Delta l$$
  7. Update the linearisation point $p_0 = \hat{p}$ and $l_0 = \hat{l}$ and iterate

Example



- Let's try to estimate a line through two points
- The points are at

$$\bar{l} = \left(\bar{l}_1\,, \bar{l}_2\,\right)^T = (1 \quad 2 \quad 3 \quad 1)^T$$

- We assume a circular co-variance $C = I$

- To be on the line parameterised by its normal vector p we require

$$g[p, l] = \begin{pmatrix} l_1^T p \\ l_2^T p \end{pmatrix} = 0$$

- To fix the length of the normal vector we need to ensure that

$$h[p] = p^T p - 1 = 0$$

Example

- We start with $p_0 = (1 \quad 0)^T$ and $\mathrm{l}_0 = \bar{l} = (1 \quad 2 \quad 3 \quad 1)^T$
- Therefore we get

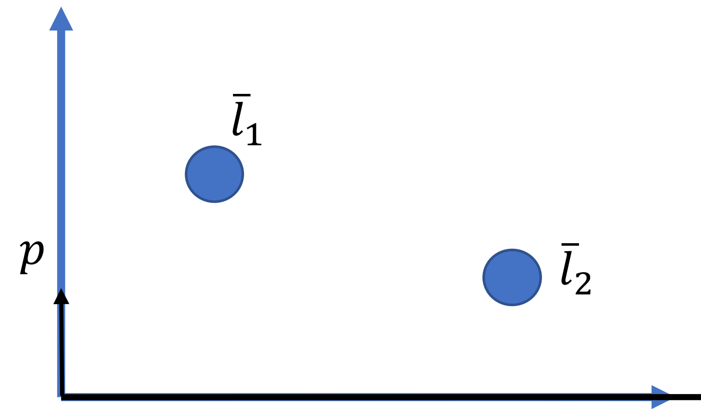
$$g_0 = g[p_0, l_0] = (1 \quad 3)^T$$
$$h_0 = h[p_0] = 0$$

- The Jacobians are

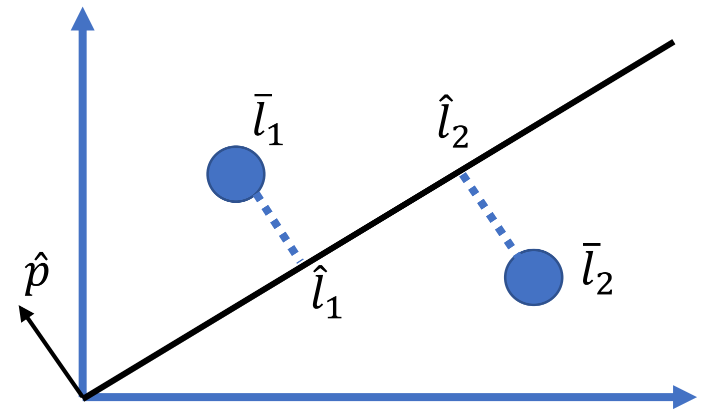
$$A = \left.\frac{\partial g}{\partial p}\right|_{p_0, l_0} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$$

$$B = \left.\frac{\partial g}{\partial l}\right|_{p_0, l_0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$H = \left.\frac{\partial h}{\partial p}\right|_{p_0, l_0} = (1 \quad 0)$$



Example



- Now we calculate

$$A^T(BCB^T)^{-1}A^T = \begin{pmatrix} 7 & 6 \\ 4 & 7 \end{pmatrix}$$

- And

$$A^T(BCB^T)^{-1}\left(B\left(l_0 - \bar{l}\right) - g_0\right) = \begin{pmatrix} -10 \\ -5 \end{pmatrix}$$

- The Normal equation system is then

$$\begin{pmatrix} 7 & 6 & 1 \\ 4 & 7 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} \Delta p \\ \mu \end{pmatrix} = \begin{pmatrix} -10 \\ -5 \\ 0 \end{pmatrix}$$

- Continuing this procedure will converge towards $\hat{p}$, $\hat{l}_1$, $\hat{l}_2$

Summary

- The Gauss-Helmert-Model is applicable when we understand the implicit relation between observable quantities and model parameters
- It is a gradient-based iterative search using the derivatives of the model functions, which therefore need to be differentiable
- Also, as all gradient-based search methods it needs an initial "guess" of the parameters to start the search and can yield local minima of the objective function

Thank you for your attention!