Linear Inverse Problems: A discrete presentation

Ali Mohammad-Djafari

A Graduated Cours

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46555, USA

Draft: May 1, 2001 *

*File: `ip1.tex`

Contents

1 Introduction and some examples

1.1 Some classical inverses problems

- 1D prblems:
  - Estimate $p(x)$ from its geometric moments

$$m_k = \int x^k p(x)\,\mathrm{d}x, \quad k = 0, \cdots, K-1$$

  - Estimate $p(x)$ from its harmonic moments

$$m_k = \int \exp\left[j2\pi k/N)\right] p(x)\,\mathrm{d}x, \quad k = 0, \cdots, K-1, K \leq N$$

- 2D prblems:
  - Estimate $p(x,y)$ from its geometric moments

$$m_{kl} = \iint x^k\, y^l p(x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad \begin{array}{l} k = 0, \cdots, K-1, \\ l = 0, \cdots, L-1 \end{array}$$

  - Estimate $p(x,y)$ from its harmonic moments

$$m_{kl} = \iint \exp\left[-j2\pi(k/N + l/N\right] p(x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad \begin{array}{l} k = 0, \cdots, K-1, \\ l = 0, \cdots, L-1 \end{array}$$

  - Estimate $p(x,y)$ from its marginals

$$p(x) = \int p(x,y)\,\mathrm{d}y \quad \text{and} \quad p(y) = \int p(x,y)\,\mathrm{d}x$$

  - Estimate $p(x,y)$ from its line integrals (Radon transform)

$$p_\phi(r) = \int_{L_{r,\phi}} p(x,y)\,\mathrm{d}l$$

- 2D shape estimation from moments:
  Assume $p(x,y) = \left\{ \begin{array}{ll} 1 & (x,y) \in \mathcal{C} \\ 0 & \text{elsewhere} \end{array} \right.$.
  - Estimate the shape $\mathcal{C}$ from the geometric moments of $p(x,y)$

$$\begin{aligned} m_{kl} &= \iint x^k\, y^l p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{(x,y)\in\mathcal{C}} x^k\, y^l\,\mathrm{d}x\,\mathrm{d}y, \quad \begin{array}{l} k = 0, \cdots, K-1, \\ l = 0, \cdots, L-1 \end{array} \end{aligned}$$

- Estimate the shape $\mathcal{C}$ from the harmonic moments of $p(x,y)$

$$\begin{aligned} m_{kl} &= \iint \exp\left[-j2\pi(k/N + l/N\right] p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{(x,y)\in\mathcal{C}} \exp\left[-j2\pi(k/N + l/N\right]\,\mathrm{d}x\,\mathrm{d}y, \qquad \begin{array}{l} k=0,\cdots,K-1, \\ l=0,\cdots,L-1 \end{array} \end{aligned}$$

- 3D problems:
  - Estimate $p(x,y,z)$ from its geometric moments

$$m_{klm} = \iiint x^k\, y^l\, z^l p(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, \qquad \begin{array}{l} k=0,\cdots,K-1, \\ l=0,\cdots,L-1, \\ m=0,\cdots,M-1 \end{array}$$

  - Estimate $p(x,y,z)$ from its harmonic moments

$$m_{klm} = \iiint \exp\left[-j2\pi(k/N + l/N + m/N\right] p(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, \qquad \begin{array}{l} k=0,\cdots,K-1, \\ l=0,\cdots,L-1, \\ m=0,\cdots,M-1 \end{array}$$

  - Estimate $p(x,y,z)$ from its 2D marginals

$$\begin{aligned} p(x,y) &= \int p(x,y,z)\,\mathrm{d}z \\ p(x,z) &= \int p(x,y,z)\,\mathrm{d}y \\ p(y,z) &= \int p(x,y,z)\,\mathrm{d}y \end{aligned}$$

  - Estimate $p(x,y,z)$ from its 1D marginals

$$\begin{aligned} p(x) &= \iint p(x,y,z)\,\mathrm{d}y\,\mathrm{d}z \\ p(y) &= \iint p(x,y,z)\,\mathrm{d}x\,\mathrm{d}z \\ p(z) &= \iint p(x,y,z)\,\mathrm{d}x\,\mathrm{d}y \end{aligned}$$

  - Estimate $p(x,y,z)$ from its 2D radiographes (X-ray transform)

$$p_\phi(u,v) = \int_{L_{u,v,\phi}} p(x,y,z)\,\mathrm{d}l$$

  - Estimate $p(x,y,z)$ from its Radon transform

$$p_{\vec{u}}(r) = \int_{S_{\vec{u},r}} p(x,y,z)\,\mathrm{d}s$$

- 3D shape estimation from moments Assume $p(x,y,z) = \begin{cases} 1 & (x,y,z) \in \mathcal{S} \\ 0 & \text{elsewhere} \end{cases}$.

  - Estimate the shape $\mathcal{S}$ from the geometric moments of $p(x,y,z)$

$$\begin{aligned} m_{klm} &= \iiint x^k\, y^l\, z^l p(x,y,z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \\ &= \iiint_{(x,y,z)\in\mathcal{S}} x^k\, y^l\, z^l\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \quad \begin{array}{l} k=0,\cdots,K-1, \\ l=0,\cdots,L-1, \\ m=0,\cdots,M-1 \end{array} \end{aligned}$$

  - Estimate the shape $\mathcal{C}$ from the harmonic moments of $p(x,y,z)$

$$\begin{aligned} m_{klm} &= \iiint \exp\left[-j2\pi(k/N + l/N + m/N\right] p(x,y,z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \\ &= \iiint_{(x,y,z)\in\mathcal{S}} \exp\left[-j2\pi(k/N + l/N + m/N\right]\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z, \quad \begin{array}{l} k=0,\cdots,K-1, \\ l=0,\cdots,L-1, \\ m=0,\cdots,M-1 \end{array} \end{aligned}$$

  - Estimate the shape $\mathcal{C}$ from its 2D marginals

$$\begin{aligned} p(x,y) &= \int_{\mathcal{C}} \mathrm{d}z \\ p(x,z) &= \int_{\mathcal{C}} \mathrm{d}y \\ p(y,z) &= \int_{\mathcal{C}} \mathrm{d}y \end{aligned}$$

  - Estimate the shape $\mathcal{C}$ from its 1D marginals

$$\begin{aligned} p(x) &= \iint_{\mathcal{C}} \mathrm{d}y\, \mathrm{d}z \\ p(y) &= \iint_{\mathcal{C}} \mathrm{d}x\, \mathrm{d}z \\ p(z) &= \iint_{\mathcal{C}} \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

  - Estimate $\mathcal{C}$ from its 2D radiographes (X-ray transform)

$$p_\phi(u,v) = \int_{L_{u,v,\phi}\cap\mathcal{C}} \mathrm{d}l$$

  - Estimate $p(x,y,z)$ from its Radon transform

$$p_{\vec{u}}(r) = \int_{S_{\vec{u},r}\cap\mathcal{C}} \mathrm{d}s$$

1.2 Deconvolution

1.3 Image restoration

1.4 Fluoroscopy

1.5 Rheology

1.6 Inverse gravitational potential

1.7 Backward heat equation

1.8 Tomography in layered media

2 Viewpoint 1: Estimating a solution

2.1 Introduction

The simplest model for a discretized linear inverse problem is

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{x}$, a vector of dimension N, represents the unknown parameters, $\boldsymbol{A}$, a matrix of dimensions $(M \times N)$, represents the model and $\boldsymbol{y}$ the data.

If $M = N$ and $\boldsymbol{A}$ is inversible, then a unique solution exists: $\widehat{\boldsymbol{x}} = \boldsymbol{A}^{-1}\boldsymbol{y}$. However, even in this case, it is not sure that we would be satisfied with this solution. To see this, consider the case where $\boldsymbol{A}$ is ill-conditionned which is very often the case in inverse problems. Then, small variations in the data, may give solutions which are very far from each other. This means that, in practical situations, where the data are contaminated by the errors, we can not really make any confidence to the computed solution.

As we will see, in all cases, it will be better to find an approximate solution, less sensitive to the errors than an exact solution, but too sensitive to the errors.

In what follows, we consider more general cases where the model (1) is under-, even- or over-determined. By under-determined we mean the case where the equation (1) has more than one solution, by even-determined the case where it has only one solution, and by overdetermined the case where it has not any exact solution.

In the following, we define $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}$ to represent the errors and $E = \boldsymbol{e}^t\boldsymbol{e}$ to represent its energy.

2.2 Overdetermined model

By *overdetermined* case we mean the situation where the equation (1) has not an exact solution. This may happen when, for example, there are more data than the parameters (The matrix $\boldsymbol{A}$ has more raws M than columns N).

Then, we have to define a criterion to select a "best" approximate solution. One of such criteria is the *least squares (LS)*. The LS solution is defined as

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{ E = \boldsymbol{e}^t\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) \right\} \tag{2}$$

The solution is obtained easily

$$\nabla_{\boldsymbol{x}} E = 0 \longrightarrow \widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{y} \tag{3}$$

Now assume that there is a "true solution" $\boldsymbol{x}^*$ such that $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{y}$. Then,

$$\widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{y} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{x}^* \tag{4}$$

which means that, in this case, the LS squares solution will find that true solution without any ambiguity. In other words, this estimate is *unbiased* (the estimation error $\widehat{\boldsymbol{x}} - \boldsymbol{x} = \boldsymbol{0}$).

However, as mentionned before, this solution may not give satisfaction in some practical applications. To see this, let consider the variability of this solution with respect to the variability of the data (errors). A measure of the variability is the covariance. We may then consider $\text{Cov}\{\boldsymbol{x}\}$ as a function of $\text{Cov}\{\boldsymbol{y}\}$:

$$\text{Cov}\{\boldsymbol{x}\} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\text{Cov}\{\boldsymbol{y}\}\,\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-t} \tag{5}$$

Assume now that $\text{Cov}\{\boldsymbol{y}\} = \boldsymbol{I}$, thus its size $\|\text{Cov}\{\boldsymbol{y}\}\|^2 = M$. Then $\text{Cov}\{\boldsymbol{x}\} = (\boldsymbol{A}^t\boldsymbol{A})^{-t}$ and if the matrix $\boldsymbol{A}^t\boldsymbol{A}$ is ill-conditionned, we may have $\|\text{Cov}\{\boldsymbol{x}\}\|^2 >> M$.

2.3 Underdetermined model

By *underdetermined* we mean that the equation has more than one solution. One sample example is the case where $M < N$ and the data are consistent. However, note that, the model (1) may be even- or over-determined even if $M < N$ when the data are not consistent.

Let assume that (1) is underdetermined. Then, we have to define a criterion to select one of these solutions to be the "best" in some sense. One of such criteria is the *minimum norm* or the *minimum length*.

The minimum norm (MN) solution is defined as:

$$\min \quad \frac{1}{2}(\boldsymbol{x}^t\boldsymbol{x}) \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{6}$$

Using the Lagrange multiplier technic, we have

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}(\boldsymbol{x}^t\boldsymbol{x}) + \sum_i \lambda_i(y_i - [\boldsymbol{A}\boldsymbol{x}]_i) = \frac{1}{2}(\boldsymbol{x}^t\boldsymbol{x}) + \boldsymbol{\lambda}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) \tag{7}$$

The optimum solution is the stationnary point of $L(\boldsymbol{x}, \boldsymbol{\lambda})$:

$$\begin{cases} \nabla_{\boldsymbol{x}} L &= \boldsymbol{x} - \boldsymbol{A}^t\lambda = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}} L &= \boldsymbol{A}\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{0} \end{cases} \tag{8}$$

We can note that $(\boldsymbol{x}, \boldsymbol{\lambda})$ is the solution of the following system of equations

$$\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{A}^t \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{y} \end{pmatrix} \tag{9}$$

This system has explicit solutions for $\boldsymbol{\lambda}$ and $\boldsymbol{x}$ which is

$$\begin{cases} \widehat{\boldsymbol{x}} &= \boldsymbol{A}^t\widehat{\boldsymbol{\lambda}} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}\boldsymbol{y} \\ \widehat{\boldsymbol{\lambda}} &= (\boldsymbol{A}\boldsymbol{A}^t)^{-1}\boldsymbol{y} \end{cases} \tag{10}$$

When we computed $\widehat{\boldsymbol{x}}$ we can calculate

$$\widehat{\boldsymbol{y}} = \boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}\boldsymbol{y} = \boldsymbol{y} \tag{11}$$

This means that, in this case, the prediction error of the data is zero ($\widehat{\boldsymbol{y}} = \boldsymbol{y}$).

Again here, this solution also may not give satisfaction in some practical applications, because in this case we have

$$\text{Cov}\{\boldsymbol{x}\} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}\text{Cov}\{\boldsymbol{y}\}\,(\boldsymbol{A}^t\boldsymbol{A})^{-t})\boldsymbol{A} \tag{12}$$

and for the case $\text{Cov}\{\boldsymbol{y}\} = \boldsymbol{I}$, we have

$$\text{Cov}\{\boldsymbol{x}\} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-2}\boldsymbol{A} \tag{13}$$

and if the matrix $\boldsymbol{A}\boldsymbol{A}^t$ is ill-conditionned, we may have $\|\text{Cov}\{\boldsymbol{x}\}\|^2 >> M$.

2.4 General case

Most inverse problems that arise in practice are neither completely overdetermined nor completely underdetermined. More precisely, we may not know *a priori* if the model is under-, even- or over-determined.

As we mentionned, the model (1) may be underdetermined even if the matrix $\boldsymbol{A}$ has more raws than columns and *vice versa.*

Ideally, we would like to sort the unknown parameters into two groups; those that are overdetermined and those that are underdetermined. One way to do this is to form a new set of parameters $\widetilde{\boldsymbol{x}}$ by linearly combining the real parameters $\boldsymbol{x}$ such that

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \longrightarrow \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{x}} = \widetilde{\boldsymbol{y}} \tag{14}$$

and such that the new parameters $\widetilde{\boldsymbol{x}}$ can be partionned into an overdetermined part $\widetilde{\boldsymbol{x}}_o$ and an underdetermined part $\widetilde{\boldsymbol{x}}_u$

$$\begin{pmatrix} \widetilde{\boldsymbol{A}}_o & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{A}}_u \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{x}}_o \\ \widetilde{\boldsymbol{x}}_u \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{y}}_o \\ \widetilde{\boldsymbol{y}}_u \end{pmatrix} \tag{15}$$

If this can be acheived, then we could determine $\widetilde{\boldsymbol{x}}_u$ in the LS sense and $\widetilde{\boldsymbol{x}}_u$ in the MN sense.

One way to do this partionning is the *singular value decomposition.* We will describe it a little later. Another way to do this in an implicit way, is to minimize some combination of the prediction error (LS) and solution norm

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{ \boldsymbol{e}^t\boldsymbol{e} + \alpha\boldsymbol{x}^t\boldsymbol{x} \right\} \tag{16}$$

As we will see below, this solution can be interpreted, either as

$$\text{minimize} \quad \boldsymbol{x}^t\boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{e}^t\boldsymbol{e} = \epsilon_e^2 \tag{17}$$

or as

$$\text{minimize} \quad \boldsymbol{e}^t\boldsymbol{e} \quad \text{s.t.} \quad \boldsymbol{x}^t\boldsymbol{x} = \epsilon_x^2 \tag{18}$$

In the first case the solution is given by the stationary point of

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{x}^t\boldsymbol{x} + \lambda[(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) - \epsilon_e^2] \tag{19}$$

which is obtained by

$$\nabla_{\boldsymbol{x}} L = 2\boldsymbol{x} - 2\lambda\boldsymbol{A}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0} \longrightarrow \widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{A} + \frac{1}{\lambda}\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{y} \tag{20}$$

λ has to be chosen in such a way that $\boldsymbol{e}^t\boldsymbol{e} = \epsilon_e^2$. Unfortunately, it is not possible to obtain an explicit expression for λ as a function of ϵ_e^2.

In the second case the solution is given by the stationary point of

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda(\boldsymbol{x}^t\boldsymbol{x} - \epsilon_x^2) \tag{21}$$

which is obtained by

$$\nabla_{\boldsymbol{x}} L = 2\boldsymbol{A}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) - 2\lambda\boldsymbol{x} = \boldsymbol{0} \longrightarrow \widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{y} \tag{22}$$

λ has to be chosen in such a way that $\boldsymbol{x}^t\boldsymbol{x} = \epsilon_x^2$.

We can remark that in both cases we obtain the same form

$$\widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{A} + \alpha\boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{y} \tag{23}$$

for the solution.

We can extend all these cases by replacing the norms by weighted norms.

2.5 Weighted LS solution

The weighted least squares (WLS) solution is defined as

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{\boldsymbol{e}^t\boldsymbol{W}_e\boldsymbol{e}\right\} \tag{24}$$

where $\boldsymbol{W}_e$ is a weighting matrix, which in general is choosed to be a symetric positive definite matrix. The solution is then given by

$$\widehat{\boldsymbol{x}} = (\boldsymbol{A}^t\boldsymbol{W}_e\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{W}_e\boldsymbol{y} \tag{25}$$

2.6 Minimum weighted norm solution

In the same way, we may generalize the MN solution to minimum weighted norme (MWN), by defining the solution as

$$\text{minimize} \quad \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t\boldsymbol{W}_x(\boldsymbol{x}-\boldsymbol{x}_0) \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{26}$$

Using again the Lagrange multiplier technic, we have

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{\lambda}) &= \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t\boldsymbol{W}_x(\boldsymbol{x}-\boldsymbol{x}_0) + \sum_i \lambda_i(y_i - [\boldsymbol{A}\boldsymbol{x}]_i) \\ &= \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t\boldsymbol{W}_x(\boldsymbol{x}-\boldsymbol{x}_0) + \boldsymbol{\lambda}^t(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) \end{aligned} \tag{27}$$

and the solution satisfies

$$\left\{\begin{array}{lcl} \nabla_{\boldsymbol{x}}L &=& \boldsymbol{W}_x(\boldsymbol{x}-\boldsymbol{x}_0) - \boldsymbol{A}^t\boldsymbol{\lambda} = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}}L &=& \boldsymbol{A}\boldsymbol{x}-\boldsymbol{y} = \boldsymbol{0} \end{array}\right. \tag{28}$$

which can be written in the following matrix form:

$$\begin{pmatrix} \boldsymbol{W}_x & -\boldsymbol{A}^t \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}-\boldsymbol{x}_0 \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0 \end{pmatrix}. \tag{29}$$

This system has explicit solution:

$$\left\{\begin{array}{lcl} \widehat{\boldsymbol{x}} &=& \boldsymbol{x}_0 + \boldsymbol{W}_x^{-1}\boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{W}_x^{-1}\boldsymbol{A}^t)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0) \\ \boldsymbol{\lambda} &=& (\boldsymbol{A}\boldsymbol{W}_x^{-1}\boldsymbol{A}^t)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0) \end{array}\right. \tag{30}$$

2.7 Minimum norme WLS

By extension, in a more general case we may define the solution as

$$\text{minimize} \quad (\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{W}_x (\boldsymbol{x}-\boldsymbol{x}_0) \quad \text{s.t.} \quad \boldsymbol{e}^t \boldsymbol{W}_e \boldsymbol{e} = \epsilon_e^2 \tag{31}$$

or as

$$\text{minimize} \quad \boldsymbol{e}^t \boldsymbol{W}_e \boldsymbol{e} \quad \text{s.t.} \quad (\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{W}_x (\boldsymbol{x}-\boldsymbol{x}_0) = \epsilon_x^2 \tag{32}$$

In the first case the Langragian is

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = (\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{W}_x (\boldsymbol{x}-\boldsymbol{x}_0) + \lambda[(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^t \boldsymbol{W}_e (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) - \epsilon_e^2] \tag{33}$$

and the solution has to satisfy

$$\nabla_{\boldsymbol{x}} L = 2\boldsymbol{W}_x(\boldsymbol{x}-\boldsymbol{x}_0) - 2\lambda \boldsymbol{A}^t \boldsymbol{W}_e (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0} \tag{34}$$

and is given by

$$\widehat{\boldsymbol{x}} = \boldsymbol{x}_0 + (\boldsymbol{A}^t \boldsymbol{W}_e \boldsymbol{A} + \lambda \boldsymbol{W}_x)^{-1} \boldsymbol{A}^t \boldsymbol{W}_e (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) \tag{35}$$

In the second case the Langragian is

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^t \boldsymbol{W}_e (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + \lambda((\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{W}_x (\boldsymbol{x}-\boldsymbol{x}_0) - \epsilon_x^2) \tag{36}$$

and the solution satisfies

$$\nabla_{\boldsymbol{x}} L = \boldsymbol{W}_e \boldsymbol{A}^t (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + 2\lambda \boldsymbol{W}_x (\boldsymbol{x}-\boldsymbol{x}_0) = \boldsymbol{0} \tag{37}$$

and is given by

$$\widehat{\boldsymbol{x}} = \boldsymbol{x}_0 + \boldsymbol{W}_x^{-1} \boldsymbol{A}^t \left(\boldsymbol{A}^t \boldsymbol{W}_x^{-1} \boldsymbol{A} + \lambda \boldsymbol{W}_e^{-1}\right)^{-1} (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) \tag{38}$$

To summarize, the solution can be written as

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}_0 + (\boldsymbol{A}^t \boldsymbol{W}_e \boldsymbol{A} + \lambda \boldsymbol{W}_x)^{-1} \boldsymbol{A}^t \boldsymbol{W}_e (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) && (39) \\ &= \boldsymbol{x}_0 + \boldsymbol{W}_x^{-1} \boldsymbol{A}^t \left(\boldsymbol{A}^t \boldsymbol{W}_x^{-1} \boldsymbol{A} + \lambda \boldsymbol{W}_e^{-1}\right)^{-1} (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) && (40) \end{aligned}$$

with the only difference that, in the first case λ has to be determined to satisfy the constraint $\boldsymbol{e}^t \boldsymbol{W}_e \boldsymbol{e} = \epsilon_e^2$ and in the second case it has to satisfy the constraint $\boldsymbol{x}^t \boldsymbol{W}_x \boldsymbol{x} = \epsilon_x^2$.

2.8 Linearly constraint WLS

minimize $\boldsymbol{e}^t \boldsymbol{W}_e \boldsymbol{e}$ subject to some constraints $\boldsymbol{F}\boldsymbol{x} - \boldsymbol{h} = \boldsymbol{0}$.

$$\text{minimize} \quad \boldsymbol{e}^t \boldsymbol{W}_e \boldsymbol{e} \quad \text{s.t.} \quad \boldsymbol{F}\boldsymbol{x} - \boldsymbol{h} = \boldsymbol{0} \tag{41}$$

Again here the solution is given by

$$\begin{pmatrix} \boldsymbol{A}^t \boldsymbol{W}_e \boldsymbol{A} & \boldsymbol{F}^t \\ \boldsymbol{F} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^t \boldsymbol{y} \\ \boldsymbol{h} \end{pmatrix} \tag{42}$$

2.9 Other criteria

Up to now, we only considered the normes $\boldsymbol{e}^t\boldsymbol{e}$, $\boldsymbol{x}^t\boldsymbol{x}$ or their weighted versions to define a solution. These are quadratic criteria with the general forme

$$Q(\boldsymbol{p},\boldsymbol{q}) = \|\boldsymbol{p}-\boldsymbol{q}\|^2_{\boldsymbol{W}} = (\boldsymbol{p}-\boldsymbol{q})^t\boldsymbol{W}(\boldsymbol{p}-qb) = \sum_i\sum_j w_{i,j}(p_i-q_i)(p_j-q_j) \tag{43}$$

with some special cases :

- $\boldsymbol{W}=\boldsymbol{D}^t\boldsymbol{D} \longrightarrow$
$$Q(\boldsymbol{p},\boldsymbol{q}) = (\boldsymbol{p}-\boldsymbol{q})^t\boldsymbol{D}^t\boldsymbol{D}(\boldsymbol{p}-qb) = \|\boldsymbol{D}(\boldsymbol{p}-\boldsymbol{q})\|^2 \tag{44}$$
- $\boldsymbol{W}=\operatorname{diag}\{w_1,\ldots,w_n\} \longrightarrow$
$$Q(\boldsymbol{p},\boldsymbol{q}) = \sum_j w_j(p_j-q_j)^2 \tag{45}$$
- $\boldsymbol{W}=\boldsymbol{I} \longrightarrow$
$$Q(\boldsymbol{p},\boldsymbol{q}) = \sum_j (p_j-q_j)^2 \tag{46}$$

$$Q(\boldsymbol{p},\boldsymbol{q}) = \|\boldsymbol{p}-\boldsymbol{q}\|^2 = (\boldsymbol{p}-\boldsymbol{q})^t(\boldsymbol{p}-qb) = \sum_j (p_j-q_j)^2 \tag{47}$$

Other criteria can be used. In particular, we can mention

- L_α norme
$$L_\alpha(\boldsymbol{p},\boldsymbol{q}) = \sum_j |p_j-q_j|^\alpha, \quad 1<\alpha<2 \tag{48}$$
- Kullback-Leibler (KL) mismatch function
$$K(\boldsymbol{p},\boldsymbol{q}) = \sum_j p_j \ln\left(\frac{p_j}{q_j}\right) - (p_j-q_j) \tag{49}$$

For example, for the over-determined case, we may define a solution as

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}\left\{K(\boldsymbol{A}\boldsymbol{x},\boldsymbol{y})\right\}. \tag{50}$$

Or, for the under-determined case, we may define the solution as

$$\text{minimize} \quad K(\boldsymbol{x},\boldsymbol{x}_0) \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{y} \tag{51}$$

In the general case, we may define a solution by

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}\left\{\Delta_1(\boldsymbol{A}\boldsymbol{x},\boldsymbol{y}) + \lambda\Delta_2(\boldsymbol{x},\boldsymbol{x}_0)\right\} \tag{52}$$

where Δ_1 and Δ_2 can be either Q in (47), L_α in (48) or K in (49) or any other distance or mismatch function.

When $J=\Delta_1+\lambda\Delta_2$ is a quadratic function of $\boldsymbol{x}$, then the solution is a linear function of the data $\boldsymbol{y}$ and it is easily computed. When J is not quadratic, but it is a strict convexe function of $\boldsymbol{x}$, the solution is no more a linear function of the data $\boldsymbol{y}$. However, it can be computed easily by any local optimization algorithme.

3 Viewpoint 2: Determining an inverse operator

3.1 Solution versus Operators

In previous chapter we derived methods for giving estimates of the possibles solutions (exact or approximate) for the linear inverse problems described by equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$. All these methods were developped by examining two properties of the solution: data prediction error and the solution simplicity (or norm). In many of those methods, the solution is obtained by a linear combination of the data $\widehat{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{y}$ where the expression of the matrix $\boldsymbol{M}$ was different for different methods.

In this chapter, we shift our attention from the solution $\widehat{\boldsymbol{x}}$ itself to the operator $\boldsymbol{M}$. Since $\boldsymbol{M}$ "solves" or "inverts" in some sense the inverse problems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$, it is often called *generalized inverse* (GI) of $\boldsymbol{A}$. We remember from the previous chapter that the GI of overdetermined LS problem is $\boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t$, and the GI of minimum norm underdetermined problem is $\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}$. Note also that for the trivial case where $\boldsymbol{A}$ is square and inversible, the both aforementioned GI matrices become the ordinary inverse $\boldsymbol{M} = \boldsymbol{A}^{-1}$.

In the following we focus the following problem:

Given that $\widehat{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{y}$, find the best matrix $\boldsymbol{M}$ in some sense in such a way that $\widehat{\boldsymbol{x}}$ solves (in some sense to be determined) the inverse problem $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$. The first thing to do then is to examine the possible and useful criteria to measure the goodness of a solution.

3.2 Parameter resolution matrix

Assume that there exists a true $\boldsymbol{x}^*$ such that $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{y}$. Then,

$$\widehat{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{y} = (\boldsymbol{M}\boldsymbol{A})\boldsymbol{x}^* = \boldsymbol{R}\boldsymbol{x}^* \tag{53}$$

The matrix $\boldsymbol{R} = \boldsymbol{M}\boldsymbol{A}$ is called the *parameter or model resolution matrix.* The ideal situation would be $\boldsymbol{R} = \boldsymbol{I}$. So a measure of goodness of the model is

$$G_1 = \|\boldsymbol{R} - \boldsymbol{I}\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(R_{i,j} - I_{i,j})^2 = \sum_{i=1}^{N} K_i \quad \text{with} \quad K_i = \sum_{j=1}^{N}(R_{i,j} - I_{i,j})^2 \tag{54}$$

We can seek for a matrix $\boldsymbol{M}$ such that G_1 be minimal.

$$G_1 = \|\boldsymbol{R} - \boldsymbol{I}\|^2 = [\boldsymbol{M}\boldsymbol{A} - \boldsymbol{I}]^t[\boldsymbol{M}\boldsymbol{A} - \boldsymbol{I}] \tag{55}$$

$$\frac{\partial G_1}{\partial \boldsymbol{M}} = [\boldsymbol{M}\boldsymbol{A} - \boldsymbol{I}]\boldsymbol{A}^t = [0] \longrightarrow \boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1} \tag{56}$$

3.3 Data resolution matrix

When we fixe the solution as $\widehat{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{y}$, we have

$$\widehat{\boldsymbol{y}} = \boldsymbol{A}\widehat{\boldsymbol{x}} = (\boldsymbol{A}\boldsymbol{M})\boldsymbol{y} = \boldsymbol{N}\boldsymbol{y} \tag{57}$$

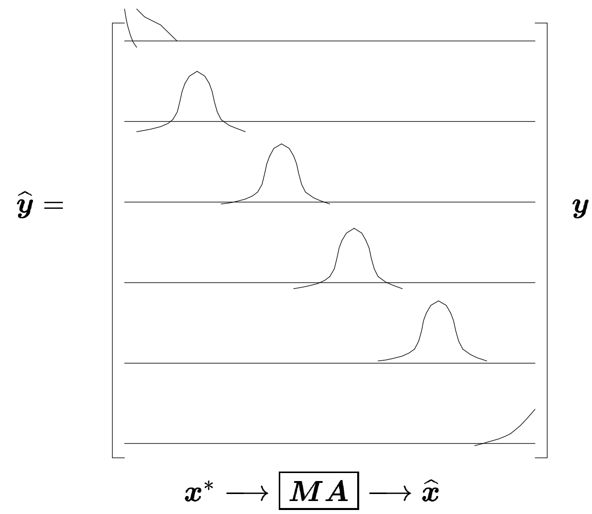

Figure 1: Model or parameter resolution matrix

The matrix $\boldsymbol{N} = \boldsymbol{A}\boldsymbol{M}$ is called the *Data or output resolution matrix.* Again the ideal situation would be $\boldsymbol{N} = \boldsymbol{I}$. So a measure of goodness of the data is is

$$G_2 = \|\boldsymbol{N} - \boldsymbol{I}\|_2^2 = \sum_{i=1}^{M}\sum_{j=1}^{M}(N_{i,j} - I_{i,j})^2 = \sum_{i=1}^{M} J_i \quad \text{with} \quad J_i = \sum_{j=1}^{M}(N_{i,j} - I_{i,j})^2 \tag{58}$$

$$\boldsymbol{y} \longrightarrow \boxed{\boldsymbol{N} = \boldsymbol{A}\boldsymbol{M}} \longrightarrow \widehat{\boldsymbol{y}}$$

Figure 2: Data resolution matrix

Here again we can seek for a matrix $\boldsymbol{M}$ such that G_2 be minimal.

$$G_2 = \|\boldsymbol{N} - \boldsymbol{I}\|^2 = [\boldsymbol{A}\boldsymbol{M} - \boldsymbol{I}]^t[\boldsymbol{A}\boldsymbol{M} - \boldsymbol{I}] \tag{59}$$

$$\frac{\partial G_2}{\partial \boldsymbol{M}} = \boldsymbol{A}^t[\boldsymbol{A}\boldsymbol{M} - \boldsymbol{I}] = [0] \longrightarrow \boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t \tag{60}$$

Remarks:

- The criteria G_1 and G_2 are called *Dirichlet spread functions.*
- A weighted version of G_1

$$S(\boldsymbol{R}) = \sum_i\sum_j W_{i,j}(R_{i,j} - I_{i,j})^2 \tag{61}$$

is called *Backus-Gilbert* criterion.

3.4 Unit covariance or error amplification matrix

Noting that

$$\mathrm{Cov}\{\widehat{\boldsymbol{x}}\} = \mathrm{Cov}\{\boldsymbol{M}\boldsymbol{y}\} = \boldsymbol{M}\,\mathrm{Cov}\{\boldsymbol{y}\}\,\boldsymbol{M}^t \tag{62}$$

the following matrix

$$\boldsymbol{U} = \boldsymbol{M}\boldsymbol{M}^t \tag{63}$$

corresponds to the covariance of the solution if the covariance of the data $\mathrm{Cov}\{\boldsymbol{y}\} = \boldsymbol{I}$. This matrix then measures, in somehow, the error amplification from the data to the solution. So, the size of the matrix $\boldsymbol{U}$

$$G_3 = \|\boldsymbol{U}\|_2^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} U_{i,j}^2 = \sum_{i=1}^{N} L_i \quad \text{with} \quad L_i = \sum_{j=1}^{N} U_{i,j}^2 \tag{64}$$

is another quantity which can be used for the design of the matrix $\boldsymbol{M}$.

3.5 A Combined criterion

It is interesting to define a criterion

$$\begin{aligned} J &= \alpha_1 G_1 + \alpha_2 G_2 + \alpha_3 G_3 \\ &= \alpha_1 \|\boldsymbol{M}\boldsymbol{A} - \boldsymbol{I}\|_2^2 + \alpha_2 \|\boldsymbol{A}\boldsymbol{M} - \boldsymbol{I}\|_2^2 + \alpha_3 \|\boldsymbol{M}\boldsymbol{M}\|_2^2 \end{aligned}$$

and try to minimize it with respect to the elements of the matrix $\boldsymbol{M}$.

Now consider the following special cases:

- $\alpha_1 = 1$, $\alpha_2 = 0$, and $\alpha_3 = 0$

$$\frac{\partial J}{\partial \boldsymbol{M}} = \alpha_1 (\boldsymbol{M}\boldsymbol{A} - \boldsymbol{I})\boldsymbol{A}^t \tag{65}$$

  and we obtain

$$\boldsymbol{M}(\boldsymbol{A}\boldsymbol{A}^t) = \boldsymbol{A}^t \quad \longrightarrow \quad \boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1} \tag{66}$$

  Then we have

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{M}\boldsymbol{A} = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-1}\boldsymbol{A} \\ \boldsymbol{N} &= \boldsymbol{A}\boldsymbol{M} = \boldsymbol{I} \\ \boldsymbol{U} &= \boldsymbol{M}\boldsymbol{M}^t = \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{A}^t)^{-2}\boldsymbol{A} \end{aligned}$$

- $\alpha_1 = 0$, $\alpha_2 = 1$, and $\alpha_3 = 0$

$$\frac{\partial J}{\partial \boldsymbol{M}} = \alpha_2 \boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{M} - \boldsymbol{I}) \tag{67}$$

  and we obtain

$$(\boldsymbol{A}^t\boldsymbol{A})\boldsymbol{M} = \boldsymbol{A}^t \quad \longrightarrow \quad \boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t \tag{68}$$

  and we have

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{M}\boldsymbol{A} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{I} \\ \boldsymbol{N} &= \boldsymbol{A}\boldsymbol{M} = \boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t \\ \boldsymbol{U} &= \boldsymbol{M}\boldsymbol{M}^t = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-t} = (\boldsymbol{A}^t\boldsymbol{A})^{-t} \end{aligned}$$

- $\alpha_1 = 1$, $\alpha_2 = 0$, and $\alpha_3 = 1$

$$\frac{\partial J}{\partial \boldsymbol{M}} = \alpha_1(\boldsymbol{MA} - \boldsymbol{I})\boldsymbol{A}^t + \alpha_3 \boldsymbol{M} \tag{69}$$

and we obtain

$$\boldsymbol{M}(\boldsymbol{AA}^t + \lambda \boldsymbol{I}) = \boldsymbol{A}^t \quad \longrightarrow \quad \boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1} \quad \text{with} \quad \lambda = \frac{\alpha_3}{\alpha_1} \tag{70}$$

Then we have

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{MA} = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}\boldsymbol{A} \\ \boldsymbol{N} &= \boldsymbol{AM} = \boldsymbol{AA}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1} \\ \boldsymbol{U} &= \boldsymbol{MM}^t = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-2}\boldsymbol{A} \end{aligned}$$

- $\alpha_1 = 0$, $\alpha_2 = 1$, and $\alpha_3 = 1$

$$\frac{\partial J}{\partial \boldsymbol{M}} = \alpha_2 \boldsymbol{A}^t(\boldsymbol{AM} - \boldsymbol{I}) + \alpha_3 \boldsymbol{M} \tag{71}$$

and we obtain

$$(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})\boldsymbol{M} = \boldsymbol{A}^t \quad \longrightarrow \quad \boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t \quad \text{with} \quad \lambda = \frac{\alpha_2}{\alpha_1} \tag{72}$$

and we have

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{MA} = (\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{A} \\ \boldsymbol{N} &= \boldsymbol{AM} = \boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t \\ \boldsymbol{U} &= \boldsymbol{MM}^t = (\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{A}(\boldsymbol{A} + \lambda \boldsymbol{I})^{-1} \end{aligned}$$

The following tables summarizes these cases

α_1	α_2	α_3	$\boldsymbol{M}$ satisfies	$\boldsymbol{M}$
1	0	0	$\boldsymbol{M}(\boldsymbol{AA}^t) = \boldsymbol{A}^t$	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t)^{-1}$
0	1	0	$(\boldsymbol{A}^t\boldsymbol{A})\boldsymbol{M} = \boldsymbol{A}^t$	$\boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t$
1	0	1	$\boldsymbol{M}(\boldsymbol{AA}^t + \lambda \boldsymbol{I}) = \boldsymbol{A}^t$	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}$
0	1	1	$(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})\boldsymbol{M} = \boldsymbol{A}^t$	$\boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t$
1	1	0	$\boldsymbol{M}(\boldsymbol{AA}^t) + [(\boldsymbol{A}^t\boldsymbol{A})\boldsymbol{M}]^t = 2\boldsymbol{A}^t$	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t)^{-1}$
1	1	1	$\boldsymbol{M}(\boldsymbol{AA}^t) + [(\boldsymbol{A}^t\boldsymbol{A})\boldsymbol{M}]^t + \boldsymbol{M} = 2\boldsymbol{A}^t$	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}$

α_1	α_2	α_3	$\boldsymbol{M}$	$\boldsymbol{N} = \boldsymbol{AM}$	$\boldsymbol{R} = \boldsymbol{MA}$	$\boldsymbol{U} = \boldsymbol{MM}^t$
1	0	0	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t)^{-1}$	$\boldsymbol{I}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t)^{-1}\boldsymbol{A}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t)^{-2}\boldsymbol{A}^t$
0	1	0	$\boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t$	$\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t$	$\boldsymbol{I}$	$(\boldsymbol{A}^t\boldsymbol{A})^{-1}$
1	0	1	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}$	$\boldsymbol{AA}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-2}\boldsymbol{A}$
0	1	1	$\boldsymbol{M} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t$	$\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t$	$(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{A}$	$(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^t\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A} + \lambda$
1	1	0	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t)^{-1}$	$\boldsymbol{I}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t)^{-1}\boldsymbol{A}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t)^{-2}\boldsymbol{A}$
1	1	1	$\boldsymbol{M} = \boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}$	$\boldsymbol{AA}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}$	$\boldsymbol{A}^t(\boldsymbol{AA}^t + \lambda \boldsymbol{I})^{-2}\boldsymbol{A}$

3.6 Backus-Gilbert

4 Viewpoint 3: A probabilistic approach

4.1 Likelihood function and ML solution

4.2 Joint distribution

4.3 Posterior distribution

4.4 Gaussian case

4.5 Unimodal and concave distributions

4.6 Multimodal distributions