Validating Image Analysis Algorithms

Questions

• Does the image analysis program in fact implement what the design indicated should have been implemented?

• Does the system behave in the expected way?

• Does the system meet its performance specification?
Validating Image Analysis Algorithms

• Model fitting methodology
  – Model parameter estimation
  – Registration
  – Alignment

• Ground truth based methodology
  – Border detection
  – Segmentation
  – Anomaly recognition

• Performance specification
  – Proper specification statement
  – Sample size for validation test
  – Validation Test
Model Fitting Methodology

In the model fitting paradigm, the unobserved ideal data vector $X$,

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

is composed of many instances of individual points $x_1, \ldots, x_N$, each associated with the model parameter vector $\Theta$.

The model states:

$$f(x_n, \Theta) = 0, \ n = 1, \ldots, N$$
Registration Example

Let
\[ u_1, \ldots, u_N \] be N points in the patient coordinate frame.
\[ v_1, \ldots, v_N \] be N corresponding points in the reference coordinate frame.
\[ \Theta \] be the transformation parameters.

If there were
- no observation noise
- no errors in point correspondence
- reality were to completely obey the transformation \( h \)

then
\[ h(v_n; \Theta) = u_n \]

In this case, \( x_n = (u_n, v_n) \) and
\[ f(x_n, \Theta) = h(v_n; \Theta) - u_n \]
Noise Model

There is a noise model that relates the observed noisy data $\hat{X}$ to the unobserved ideal $X$.

$$\hat{X} = X + \xi$$

where $\xi$ is assumed to have a mean 0 and a covariance $\Sigma$.

Just as $X$ was composed of individual points $\xi$ is also composed of individual points.

$$\hat{x}_n = x_n + \xi_n$$

where $\xi_n$ has mean 0 and covariance $\Sigma_n$ and $\xi_m$ is uncorrelated with $\xi_n, m \neq n$.

$$\Sigma = \begin{pmatrix}
\Sigma_1 & 0 & \ldots & 0 \\
0 & \Sigma_2 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & \Sigma_N
\end{pmatrix}$$
The Criterion Function

There is a non-negative function $F(X, \Theta)$ satisfying

$$F(X, \Theta) = 0$$

For a given $\hat{X}$, the image analysis algorithm determines a $\hat{\Theta}$ that minimizes $F(\hat{X}, \hat{\Theta})$.

- What is the covariance matrix $\Sigma_{\hat{\Theta}\hat{\Theta}}$?
- What should the form of $F$ be?
Model

\[ X \quad \text{Ideal Noiseless Unobserved} \]
\[ \hat{X} = X + \Delta X \quad \text{Noisy Observable} \]
\[ \Theta \quad \text{Ideal True Parameter} \]
\[ \hat{\Theta} = \Theta + \Delta \Theta \quad \text{Estimated Parameter} \]
\[ \mathcal{F} \quad \text{Functional Form} \]

Problem Statement

Suppose \( \Theta \) minimizes \( \mathcal{F}(X, \Theta) \).
Find \( \hat{\Theta} \) to minimize \( \mathcal{F}(\hat{X}, \hat{\Theta}) \).
Least Squares
\[ \mathcal{F}(\hat{X}, \hat{\Theta}) \]

Maximum Likelihood Estimation
\[ \mathcal{F}(\hat{X}, \hat{\Theta}) = -\log P(\hat{X} | \hat{\Theta}) \]

Bayesian Estimation
\[ \mathcal{F}(\hat{X}, \hat{\Theta}) = -\log P(\hat{X} | \hat{\Theta}) P(\hat{\Theta}) \]
Correlation

\[ \mathcal{F}(X, \theta) = X_\theta \ast h \]

where \( X_\theta \) is the image \( X \) translated by \( \theta \).

\[ \hat{\theta} = \arg \max X_\theta \ast h \]
Curve Fitting

\( \Psi \) \hspace{1cm} \text{Unknown Free Parameters of Curve}

\( \hat{\Psi} \) \hspace{1cm} \text{Estimated Free Parameters of Curve}

\( x_i \) \hspace{1cm} \text{Ideal Noiseless Point on Curve}

\( \hat{x}_i = x_i + \Delta x_i \) \hspace{1cm} \text{Noisy Observation}

\( f \) \hspace{1cm} \text{Form of Curve}

Problem Statement

Given that

\[
\begin{align*}
  f(x_i, \Psi) &= 0, \quad i = 1, \ldots, I \\
  h(\Psi) &= 0
\end{align*}
\]

Find \( \hat{\Psi} \) to minimize

\[
\sum_{i=1}^{I} f^2(\hat{x}_i, \hat{\Psi})
\]

subject to the constraint \( h(\hat{\Psi}) = 0 \).
Let

\[ X = (x_1, \ldots, x_I) \]
\[ \hat{X} = (\hat{x}_1, \ldots, \hat{x}_I) \]
\[ \Theta = (\Psi, 0) \]
\[ \hat{\Theta} = (\hat{\Psi}, \lambda) \]

Define

\[ \mathcal{F}(\hat{X}, \hat{\Theta}) = \sum_{i=1}^{I} f^2(\hat{x}_i, \hat{\Psi}) + \lambda h(\hat{\Psi}) \]

Find \( \hat{\Theta} \) to minimize \( \mathcal{F}(\hat{X}, \hat{\Theta}) \) where \( \Theta \) minimizes \( \mathcal{F}(X, \Theta) \).
Exterior Orientation

\((x_n, y_n, z_n)\) \(n^{th}\) 3D Model Point

\((u_n, v_n)\) \(n^{th}\) Unobserved Noiseless 2D Perspective Projection of \((x_n, y_n, z_n)\)

\((\hat{u}_n, \hat{v}_n)\) Observed Noisy 2D Perspective Projection of \((x_n, y_n, z_n)\)

\(\psi\) Unknown Rotation Parameters

\(\hat{\psi}\) Estimated Rotation Parameters

\(t\) Unknown Translation Parameters

\(\hat{t}\) Estimated Translation Parameters

**Model**

\[(u_n, v_n)' = \frac{k}{r_n} (p_n, q_n)' \text{ where} \]

\[(p_n, q_n, r_n)' = R(\psi)(x_n, y_n, z_n)' + t \]

where \(R(\psi)\) is the \(3 \times 3\) rotation matrix corresponding to the rotation angle vector \(\psi\).

\[(\hat{u}_n, \hat{v}_n) = (u_n, v_n) + (\Delta u_n, \Delta v_n)\]
Problem Statement

Let

\[ \Theta = (\psi, t) \]
\[ \hat{\Theta} = (\hat{\psi}, \hat{t}) \]
\[ X = < (x_n, y_n, z_n) : n = 1, \ldots, N; (u_n, v_n) : n = 1, \ldots, N > \]
\[ \hat{X} = < (x_n, y_n, z_n) : n = 1, \ldots, N; (\hat{u}_n, \hat{v}_n) : n = 1, \ldots, N > \]

Define

\[ F(\hat{X}, \hat{\Theta}) = \sum_{i=1}^{N} f_n(\hat{u}_n, \hat{v}_n, \hat{\psi}, \hat{t}) \]

where

\[ f_n(\hat{u}_n, \hat{v}_n, \hat{\psi}, \hat{t}) \]
\[ = [\hat{u}_n - k \frac{(1, 0, 0)(R(\hat{\psi})(x_n, y_n, z_n)' + \hat{t})}{(0, 0, 1)(R(\hat{\psi})(x_n, y_n, z_n)' + \hat{t})}]^2 \]
\[ + [\hat{v}_n - k \frac{(0, 1, 0)(R(\hat{\psi})(x_n, y_n, z_n)' + \hat{t})}{(0, 0, 1)(R(\hat{\psi})(x_n, y_n, z_n)' + \hat{t})}]^2 \]
Find $\hat{\Theta}$ to minimize $\mathcal{F}(\hat{X}, \hat{\Theta})$ where $\Theta$ minimizes $\mathcal{F}(X, \Theta)$. 
Error Propagation

How does the random perturbation $\Delta X$ acting on vector $X$ propagate to the random perturbation $\Delta \Theta$ acting on the parameter vector $\Theta$?

The solution $\hat{\Theta} = \Theta + \Delta \Theta$ minimizing $\mathcal{F}(X + \Delta X, \hat{\Theta})$, must be a zero of $g(X + \Delta X, \hat{\Theta})$, the gradient of $\mathcal{F}$.

The gradient $g$ of $\mathcal{F}$ is a $K \times 1$ vector function.

$$g(X, \Theta) = \frac{\partial \mathcal{F}}{\partial \Theta}(X, \Theta)$$
Solution

To determine the effect that $\Delta X$ has on $\Delta \Theta$, we take a first order Taylor series expansion of $g$ around $(X, \Theta)$:

\[
g^{K\times1}(X + \Delta X, \Theta + \Delta \Theta) = g^{K\times1}(X, \Theta) + \frac{\partial g^{K\times N}(X, \Theta)'}{\partial X} \Delta X^{N\times1} + \frac{\partial g^{K\times K}(X, \Theta)'}{\partial \Theta} \Delta \Theta^{K\times1}
\]

But since $\Theta + \Delta \Theta$ extremizes $\mathcal{F}(X + \Delta X, \Theta + \Delta \Theta)$

\[
g(X + \Delta X, \Theta + \Delta \Theta) = 0.
\]

Since $\Theta$ extremizes $\mathcal{F}(X, \Theta)$,

\[
g(X, \Theta) = 0.
\]

Therefore,

\[
0 = \frac{\partial g(X, \Theta)'}{\partial X} \Delta X + \frac{\partial g(X, \Theta)'}{\partial \Theta} \Delta \Theta
\]
Since the relative extremum of $\mathcal{F}$ is a relative minimum, the $K \times K$ matrix
\[
\frac{\partial g}{\partial \Theta}(X, \Theta) = \frac{\partial f^2}{\partial^2 \Theta}(X, \Theta)
\]
must be positive definite for all $(X, \Theta)$. This implies that
\[
\frac{\partial g}{\partial \Theta}(X, \Theta) = \frac{\partial f^2}{\partial^2 \Theta}(X, \Theta)
\]
is non-singular. Hence
\[
\left(\frac{\partial g}{\partial \Theta}\right)^{-1}
\]
exists and we can write:
\[
\Delta \Theta = - \left(\frac{\partial g}{\partial \Theta}(X, \Theta)\right)^{-1} \left(\frac{\partial g}{\partial X}(X, \Theta)\right)' \Delta X
\]
Covariance Matrix

\[ \Delta \Theta = - \left( \frac{\partial g}{\partial \Theta}(X, \Theta) \right)^{-1} \left( \frac{\partial g}{\partial X}(X, \Theta) \right)' \Delta X \]

If \( E[\Delta X] = 0 \), then \( E[\Delta \Theta] = 0 \).

Let \( \Sigma_{\Delta \Theta \Delta \Theta} \) be the covariance matrix of the random perturbation \( \Delta \Theta \).

\[
\Sigma_{\Delta \Theta \Delta \Theta} = E[\Delta \Theta \Delta \Theta'] = E[- \left( \frac{\partial g}{\partial \Theta} \right)^{-1} \left( \frac{\partial g}{\partial X} \right)' \Delta X (- \left( \frac{\partial g}{\partial \Theta} \right)^{-1} \left( \frac{\partial g}{\partial X} \right)' \Delta X)'] = \left( \frac{\partial g}{\partial \Theta} \right)^{-1} \left( \frac{\partial g}{\partial X} \right)' E[\Delta X \Delta X'] \left( \frac{\partial g}{\partial X} \right) \left( \frac{\partial g}{\partial \Theta} \right)^{-1} = \left( \frac{\partial g}{\partial \Theta} \right)^{-1} \left( \frac{\partial g}{\partial X} \right)' \Sigma_{\Delta X \Delta X} \left( \frac{\partial g}{\partial X} \right) \left( \frac{\partial g}{\partial \Theta} \right)^{-1} \]
Covariance Matrix

$$\Delta \Theta = - \left( \frac{\partial g}{\partial \Theta}(X, \Theta) \right)^{-1} \left( \frac{\partial g}{\partial X}(X, \Theta) \right)' \Delta X$$

$$\Delta \Theta \Delta X' = \left( \frac{\partial g}{\partial \Theta}(X, \Theta) \right)^{-1} \left( \frac{\partial g}{\partial X}(X, \Theta) \right)' \Delta X \Delta X'$$

$$E[\Delta \Theta \Delta X'] = \left( \frac{\partial g}{\partial \Theta}(X, \Theta) \right)^{-1} \left( \frac{\partial g}{\partial X}(X, \Theta) \right)' E[\Delta X \Delta X']$$

$$\Sigma_{\Delta \Theta \Delta X} = \left( \frac{\partial g}{\partial \Theta}(X, \Theta) \right)^{-1} \left( \frac{\partial g}{\partial X}(X, \Theta) \right)' \Sigma_{\Delta X \Delta X}$$
Thus to the extent that the first order approximation is good, (i.e. $E[\Delta \Theta] = 0$), then

$$\Sigma_{\hat{\Theta}\hat{\Theta}} = \Sigma_{\Delta \Theta \Delta \Theta}$$
Estimated Covariance Matrix

Expand $g(X, \Theta)$ around $g(X + \Delta X, \Theta + \Delta \Theta) = g(\hat{X}, \hat{\Theta})$.

\[
g(X, \Theta) = g(\hat{X}, \hat{\Theta}) - \left( \frac{\partial g}{\partial X}(\hat{X}, \hat{\Theta}) \right)' \Delta X - \frac{\partial g}{\partial \Theta}(\hat{X}, \hat{\Theta}) \Delta \Theta
\]

In a similar manner,

\[
\Delta \Theta = -\left( \frac{\partial g}{\partial \Theta}(\hat{X}, \hat{\Theta}) \right)^{-1} \left( \frac{\partial g}{\partial X}(\hat{X}, \hat{\Theta}) \right)' \Delta X
\]

This motivates the estimator $\hat{\Sigma}_{\Delta \Theta \Delta \Theta}$ for $\Sigma_{\Delta \Theta \Delta \Theta}$ defined by

\[
\hat{\Sigma}_{\Delta \Theta \Delta \Theta} = \left( \frac{\partial g}{\partial \Theta}(\hat{X}, \hat{\Theta}) \right)^{-1} \left( \frac{\partial g}{\partial X}(\hat{X}, \hat{\Theta}) \right)' \Sigma_{\Delta X \Delta X} \times \left( \frac{\partial g}{\partial X}(\hat{X}, \hat{\Theta}) \right) \left( \frac{\partial g}{\partial \Theta}(\hat{X}, \Theta) \right)^{-1}
\]
So to the extent that the first order approximation is good, \( \hat{\Sigma}_{\hat{\Theta} \hat{\Theta}} = \hat{\Sigma}_{\Delta \Theta \Delta \Theta} \).

The relation giving the estimate \( \hat{\Sigma}_{\hat{\Theta} \hat{\Theta}} \) in terms of the computable

\[
\frac{\partial g}{\partial \Theta}(\hat{X}, \hat{\Theta}) \text{ and } \frac{\partial g}{\partial X}(\hat{X}, \hat{\Theta})
\]

means that an estimated covariance matrix for the computed \( \hat{\Theta} \) can also be calculated at the same time that the estimate \( \hat{\Theta} \) of \( \Theta \) is calculated.
Closest Distance

Given an \( \hat{x} \) and a covariance matrix \( \Sigma_{\hat{x}\hat{x}} \), find the minimizing value of

\[
(\hat{x} - x)' \Sigma_{\hat{x}\hat{x}}^{-1} (\hat{x} - x)
\]

taken over all \( x \) satisfying

\[
h(x) = 0
\]

where we know that \( \hat{x} \) is not too far from the minimizing \( x \).
Largest Probability

If \( \hat{x} \) has a Normal distribution with mean \( x \) and covariance \( \Sigma_{\hat{x}\hat{x}} \), then the density function for \( \hat{x} \) is

\[
p(\hat{x}) = \frac{1}{(2\pi)^{N/2} | \Sigma_{\hat{x}\hat{x}} |^{1/2} \exp - \frac{1}{2}(\hat{x} - x)'\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x)}
\]

Then the log density is

\[
\log p(\hat{x}) = -\frac{N \log 2\pi + \log | \Sigma_{\hat{x}\hat{x}} |}{2} - \frac{1}{2}(\hat{x} - x)'\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x)
\]

Thus the value of \( \hat{x} \) that minimizes

\[
(\hat{x} - x)'\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x)
\]

over the constraint \( h(x) = 0 \) is the value of \( \hat{x} \) that maximizes the log density.
Closest Distance

Define

$$
\epsilon^2 = (\hat{x} - x)'\Sigma^{-1}_{\hat{x}\hat{x}}(\hat{x} - x) + \lambda h(x)
$$

Then the minimizing $x$ must satisfy

$$
\frac{\partial \epsilon^2}{\partial x} = 0
$$

Now

$$
\frac{\partial \epsilon^2}{\partial x} = 2\Sigma^{-1}_{\hat{x}\hat{x}}(\hat{x} - x)(-1) + \lambda \frac{\partial h(x)}{\partial x}
$$

Hence

$$
0 = -2\Sigma^{-1}_{\hat{x}\hat{x}}(\hat{x} - x) + \lambda \frac{\partial h(x)}{\partial x}
$$

$$
\Sigma^{-1}_{\hat{x}\hat{x}}(\hat{x} - x) = \frac{\lambda \partial h(x)}{2} \frac{\partial x}{\partial x}
$$
Closest Distance

Since $\hat{x}$ is not far from the minimizing $x$, we can write

$$h(\hat{x}) = h(x) + \frac{\partial h(x)'}{\partial x} (\hat{x} - x)$$

And to a first order approximation we assume,

$$\frac{\partial h(x)}{\partial x} = \frac{\partial h(\hat{x})}{\partial x}$$

And since $h(x) = 0$

$$h(\hat{x}) = \frac{\partial h(\hat{x})'}{\partial x} (\hat{x} - x)$$

Now, $\frac{\partial h(x)}{\partial x} = \frac{\partial h(\hat{x})}{\partial x}$ implies

$$\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x) = \frac{\lambda \partial h(\hat{x})}{2 \partial x}$$

$$\hat{x} - x = \frac{\lambda}{2} \Sigma_{\hat{x}\hat{x}} \frac{\partial h(\hat{x})}{\partial x}$$
Closest Distance

Multiply both sides by $\frac{\partial h(\hat{x})'}{\partial x}$.

$$\frac{\partial h(\hat{x})'}{\partial x}(\hat{x} - x) = \frac{\lambda \partial h(\hat{x})'}{2 \partial x} \sum \hat{x}\hat{x} \frac{\partial h(\hat{x})}{\partial x}$$

$$\frac{\lambda}{2} = \frac{\partial h(\hat{x})'}{\partial x} (\hat{x} - x) \sum \hat{x}\hat{x} \frac{\partial h(\hat{x})}{\partial x}$$

But

$$h(\hat{x}) = \frac{\partial h(\hat{x})'}{\partial x} (\hat{x} - x)$$

Hence,

$$\frac{\lambda}{2} = \frac{h(\hat{x})}{\partial x} \sum \hat{x}\hat{x} \frac{\partial h(\hat{x})}{\partial x}$$
Closest Distance

Now
\[
\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x) = \frac{\lambda \partial h(\hat{x})}{2 \partial x}
\]
\[
(\hat{x} - x)'\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x) = \frac{\lambda}{2} (\hat{x} - x)' \frac{\partial h(\hat{x})}{\partial x}
\]
\[
= \frac{\lambda}{2} h(\hat{x})
\]
\[
= h^2(\hat{x})
\]
\[
= \frac{\partial h(\hat{x})}{\partial x} \sum_{\hat{x}\hat{x}} \frac{\partial h(\hat{x})}{\partial x}
\]

Therefore, the minimizing value of

\[
(\hat{x} - x)'\Sigma_{\hat{x}\hat{x}}^{-1}(\hat{x} - x)
\]

taken over all \( x \) satisfying \( h(x) = 0 \) is

\[
\frac{h^2(\hat{x})}{\frac{\partial h(\hat{x})}{\partial x} \sum_{\hat{x}\hat{x}} \frac{\partial h(\hat{x})}{\partial x}}
\]
Finding The Minimizing $x$

Since,

$$
\frac{\lambda}{2} = \frac{h(\hat{x})}{\frac{\partial h(\hat{x})}{\partial x} \sum \hat{x} \hat{x} \frac{\partial h(\hat{x})}{\partial x}}
$$

and

$$(\hat{x} - x) = \frac{\lambda}{2} \sum \hat{x} \hat{x} \frac{\partial h(\hat{x})}{\partial x}$$

The minimizing $x$ can be computed by

$$
x = \hat{x} - \frac{h(\hat{x})}{\frac{\partial h(\hat{x})}{\partial x} \sum \hat{x} \hat{x} \frac{\partial h(\hat{x})}{\partial x}} \sum \hat{x} \hat{x} \frac{\partial h(\hat{x})}{\partial x}
$$
Covariance For Any Parameter Model Fitting Problem

In the parameter model fitting problem, the unobserved ideal $X$ is composed of many instances of individual points $x_1, \ldots, x_N$, each associated with the model parameter $\Theta$.

The model states:

$$f(x_n, \Theta) = 0, \ n = 1, \ldots, N$$
Noise Model

There is a noise model that relates the observed noisy data $\hat{X}$ to the unobserved ideal $X$.

$$\hat{X} = X + \xi$$

where $\xi$ is assumed to have a mean 0 and a covariance $\Sigma_{\xi\xi}$.

Just as $X$ was composed of individual points $\xi$ is also composed of individual points.

$$\hat{x}_n = x_n + \xi_n$$

where $\xi_n$ has mean 0 and covariance $\Sigma_{x_nx_n}$ and $\xi_m$ is uncorrelated with $\xi_n$, $m \neq n$. 
\[ \Sigma \dot{\mathbf{x}} \dot{\mathbf{x}} = \begin{pmatrix} \Sigma x_1 x_1 & 0 & \ldots & 0 \\ 0 & \Sigma x_2 x_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \Sigma x_N x_N \end{pmatrix} \]
The Criterion Function

The Criterion function for the determination of the unknown parameter $\Theta$ finds that value of $\Theta$ so that the sum of the minimizing distances, in the norm of $\Sigma^{-1}$, between the observed points and the minimizing points is minimized.

$$F(X, \Theta) = F(x_1, \ldots, x_N, \Theta)$$

$$= \sum_{n=1}^{N} f^2(x_n, \Theta) \left( \frac{\partial h(x_n)}{\partial x} \right)^T \sum_{x_n x_n} \frac{\partial h(x_n)}{\partial x}$$
Gradient of the Criterion

Function

\[ g = \frac{\partial F}{\partial \Theta} \]

\[ = \frac{1}{N} \sum_{n=1}^{N} \partial \sum_{n=1}^{N} \partial x \partial f(x_n, \Theta) \partial \hat{x}_n \hat{x}_n \partial f(x_n, \Theta) \]

\[ = \frac{1}{N} \sum_{n=1}^{N} f^2(x_n, \Theta) \partial \left( \frac{\partial f}{\partial \Theta} (x_n, \Theta) \right) \]

\[ = \frac{1}{N} \sum_{n=1}^{N} 2f(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta) \]

\[ - \frac{1}{N} \sum_{n=1}^{N} f^2(x_n, \Theta) \partial \left( \frac{\partial f}{\partial \Theta} (x_n, \Theta) \right) \]

\[ = \frac{1}{N} \sum_{n=1}^{N} 2f(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta) \]

\[ - \frac{f^2(x_n, \Theta) \partial \left( \frac{\partial f}{\partial \Theta} (x_n, \Theta) \right)}{\left( \frac{\partial f}{\partial \Theta} (x_n, \Theta) \right)^2} \]
Taking Partial Derivatives
Of A Product Of A Scaler
Function With a Vector
Function

Suppose that $f$ is a scaler function of a $K \times 1$ vector variable $\Theta$ and that $v$ is a $M \times 1$ vector function of a vector variable $\Theta$. Then,

$$\frac{\partial}{\partial \Theta} f(\Theta)v(\Theta)$$

is a $K \times M$ matrix defined by

$$\frac{\partial}{\partial \Theta} f(\Theta)v(\Theta) = f(\Theta)\frac{\partial}{\partial \Theta} v(\Theta)' + \frac{\partial f}{\partial \Theta} v(\Theta)'$$
\[
\frac{\partial g}{\partial \Theta} = \sum_{n=1}^{N} 2f(x_n, \Theta) \frac{\partial}{\partial \Theta} \frac{\partial f}{\partial x}(x_n, \Theta) \left( \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x, \Theta) \right) + \\
2 \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \frac{\partial f}{\partial x}(x_n, \Theta) \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x, \Theta) \right)' - \\
f^2(x_n, \Theta) \frac{\partial}{\partial \Theta} \frac{\partial}{\partial \Theta} \left( \frac{\partial f}{\partial x}(x_n, \Theta) \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta) \right)^2 - \\
2f(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \frac{\partial}{\partial \Theta} \left( \frac{\partial f}{\partial x}(x_n, \Theta) \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta) \right) \right)'
\]
Partial of Gradient

We evaluate the partial derivative at \((x_1, \ldots, x_N)\) where

\[ f(x_n, \Theta) = 0, \quad n = 1, \ldots, N \]

Therefore,

\[
\frac{\partial g}{\partial \Theta} = 2 \sum_{n=1}^{N} \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \frac{\partial f}{\partial x}(x_n, \Theta)' \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta) \right)'
\]

\[
= 2 \sum_{n=1}^{N} \frac{\partial f}{\partial \Theta}(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta)' \left( \frac{1}{\frac{\partial f}{\partial x}(x_n, \Theta)' \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta)} \right)
\]

\[
= 2 \sum_{n=1}^{N} \frac{\partial f}{\partial \Theta}(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta)' \frac{\partial f}{\partial x}(x_n, \Theta)' \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta)
\]
Partial of Gradient

\[
\frac{\partial g}{\partial x_n} = 2f(x_n, \Theta) \frac{\partial}{\partial x_n} \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x, \Theta) \right) + \\
2 \frac{\partial f}{\partial x_n}(x_n, \Theta) \left( \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x, \Theta) \right) \right)' - \\
f^2(x_n, \Theta) \frac{\partial}{\partial x_n} \left( \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x, \Theta) \right) \right) - \\
2f(x_n, \Theta) \frac{\partial f}{\partial x_n}(x_n, \Theta) \left( \frac{\partial}{\partial \Theta} \left( \frac{\partial f}{\partial \Theta}(x_n, \Theta) \left( \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x, \Theta) \right) \right) \right)'
\]
Partial of Gradient

We evaluate the partial derivative at $(x_1, \ldots, x_N)$ where

$$f(x_n, \Theta) = 0, \; n = 1, \ldots, N$$

Therefore,

$$\frac{\partial g}{\partial x_n} = 2 \frac{\partial f}{\partial x_n}(x_n, \Theta) \left( \frac{\frac{\partial f}{\partial \Theta}(x_n, \Theta)}{\frac{\partial f}{\partial x}(x_n, \Theta) \Sigma \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta)} \right)'$$

$$= 2 \frac{\partial f}{\partial x_n}(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta)' \left( \frac{1}{\frac{\partial f}{\partial x}(x_n, \Theta) \Sigma \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta)} \right)$$
Partial of Gradient

\[
\frac{\partial g}{\partial X}^{MN \times K} = \begin{pmatrix}
    \frac{\partial g}{\partial x_1}^{M \times K} \\
    \frac{\partial g}{\partial x_2}^{M \times K} \\
    \vdots \\
    \frac{\partial g}{\partial x_N}^{M \times K}
\end{pmatrix}
\]

\[
\frac{\partial g'}{\partial X} = \begin{pmatrix}
    \frac{\partial g'}{\partial x_1} \\
    \frac{\partial g'}{\partial x_2} \\
    \vdots \\
    \frac{\partial g'}{\partial x_N}
\end{pmatrix}
= 2 \begin{pmatrix}
    \frac{\partial f}{\partial \Theta}(x_1, \Theta) \frac{\partial f}{\partial x_1}'(x_1, \Theta) \\
    \frac{\partial f}{\partial x_1}(x_1, \Theta)' \sum_{x_1} \frac{\partial f}{\partial x_1}(x_1, \Theta) \\
    \vdots \\
    \frac{\partial f}{\partial x_N}(x_N, \Theta)' \sum_{x_N} \frac{\partial f}{\partial x_N}(x_N, \Theta)
\end{pmatrix}
\]
General Case Covariance

\[
\frac{\partial g'}{\partial X} \sum \hat{x} \hat{x} \frac{\partial g}{\partial X} = 4 \sum_{n=1}^{N} \frac{\frac{\partial f}{\partial \Theta}(x_n, \Theta) \frac{\partial f}{\partial x_n}(x_n, \Theta)'}{\sum \hat{x}_n \hat{x}_n} \frac{\partial f}{\partial \Theta}(x_n, \Theta) \frac{\partial f}{\partial x_n}(x_n, \Theta)'}{\left[ \frac{\partial f}{\partial x_n}(x_n, \Theta) \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x_n}(x_n, \Theta) \right]^2}
\]

\[
= 4 \sum_{n=1}^{N} \frac{\frac{\partial f}{\partial \Theta}(x_n, \Theta) \frac{\partial f}{\partial x}(x_n, \Theta)'}{\sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta)}
\]

\[
= 2 \frac{\partial g}{\partial \Theta}(X, \Theta)
\]
General Case Covariance

$$\Sigma_{\hat{\theta} \hat{\theta}} = \left( \frac{\partial g}{\partial \Theta} \right)^{-1} \frac{\partial g}{\partial X} \sum \hat{x} \hat{x} \frac{\partial g}{\partial X} \left( \frac{\partial g}{\partial \Theta} \right)^{-1}$$

$$= \left( \frac{\partial g}{\partial \Theta} \right)^{-1} 2 \frac{\partial g}{\partial \Theta} \left( \frac{\partial g}{\partial \Theta} \right)^{-1}$$

$$= 2 \left( \frac{\partial g}{\partial \Theta} \right)^{-1}$$

$$= 2 \left( \frac{1}{2} \sum_{n=1}^{N} \frac{\partial f}{\partial x}(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta) \sum \hat{x}_n \hat{x}_n \frac{\partial f}{\partial x}(x_n, \Theta) \right)^{-1}$$

$$= \left( \frac{1}{N} \sum_{n=1}^{N} \frac{\partial f}{\partial \Theta}(x_n, \Theta) \frac{\partial f}{\partial \Theta}(x_n, \Theta) \right)^{-1}$$
General Case Covariance

\[ \frac{\partial g'}{\partial X} \hat{X} \hat{X} = 2 \sum_{n=1}^{N} \frac{\partial f(x_n, \Theta)}{\partial x_n} (x_n, \Theta)' \sum \hat{x}_n \hat{x}_n \frac{\partial f(x_n, \Theta)}{\partial x_n} (x_n, \Theta) \]

\[ \frac{\partial g}{\partial \Theta} = 2 \sum_{i=1}^{N} \frac{\partial f(x_i, \Theta)}{\partial x_i} (x_i, \Theta)' \sum \hat{x}_i \hat{x}_i \frac{\partial f(x_i, \Theta)}{\partial x_i} (x_i, \Theta) \]

\[ \Sigma \hat{\Theta} \hat{X} = - \frac{\partial g}{\partial \Theta} \frac{\partial g'}{\partial X} \sum \hat{X} \hat{X} \]

\[ = - \left[ 2 \sum_{i=1}^{N} \frac{\partial f(x_i, \Theta)}{\partial x_i} (x_i, \Theta)' \sum \hat{x}_i \hat{x}_i \frac{\partial f(x_i, \Theta)}{\partial x_i} (x_i, \Theta) \right]^{-1} \times \]

\[ = - \left[ 2 \sum_{n=1}^{N} \frac{\partial f(x_n, \Theta)}{\partial x_n} (x_n, \Theta)' \sum \hat{x}_n \hat{x}_n \frac{\partial f(x_n, \Theta)}{\partial x_n} (x_n, \Theta) \right]^{-1} \times \]
Hypothesis Testing

\[ \Sigma = \Sigma_0 \text{ and } \mu = \mu_0 \]

Define:
\[ \bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n \]

and
\[ S = \frac{1}{N - 1} \sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})' \]

where the data vectors \( x_n \) are \( p \)-dimensional and the sample size is \( N \).

Define
\[ B = (N - 1)S \]

and
\[ \lambda = (e/N)^{pN/2} |B\Sigma_0^{-1}|^{N/2} \times \exp \left( -\left[ tr(B\Sigma_0^{-1}) + N(\bar{x} - \mu_0)'\Sigma_0^{-1}(\bar{x} - \mu_0) \right]/2 \right) \]

Test statistic:
\[ T = -2 \log \lambda \]
Distribution under true null hypothesis is Chi-squared:

\[ T \sim \chi^2_{p(p+1)/2+p} \]

Reference: Anderson page 442.