Covariance, Gauge Freedom and All That

Bill Triggs

INRIA Rhône-Alpes
Grenoble, France

http://www.inrialpes.fr/movi/people/Triggs

Contents

Gauge Freedom

- Many geometric estimation problems have an internal **gauge freedom**

— an arbitrary choice of internal coordinates is required to formulate the problem

— *e.g.* the arbitrary choice of 3D coordinates in SFM

- Gauge freedom is usually associated with an internal symmetry or **gauge transformation group**

— *e.g.* 3D Euclidean, projective groups.

Problems with Gauge Freedom

- In principle, we can use whichever gauge we like

— measurable quantites like image projections do not depend on the gauge

— reconstructed results depend only 'trivially' on it : results in different gauges are equivalent modulo the gauge group.

- But in practice gauge freedom has some subtle / problematic consequences

— nontrivial topology of error surface & solution space

— effect on numerical methods, rank deficiency

— covariances and gauge uncertainty

Topology of Error Surface

- Gauge transformations act on parameter space, moving points along **gauge orbits**

— **gauge directions** are the directions locally tangent to the orbits

- All parameter vectors on the same orbit are equivalent modulo a gauge transformation, and hence have the same image projections and residual error.

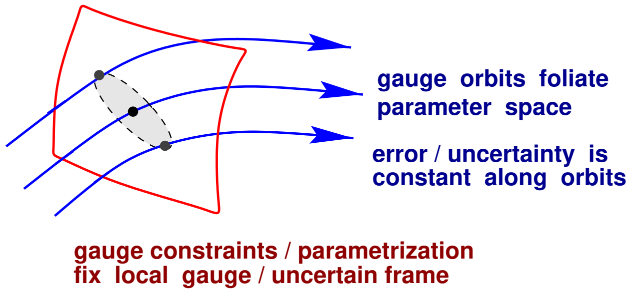
- So a gauged error minimization problem does not have a unique solution, but rather an equivalence class modulo gauge.

Gauge Fixing

- For fixed, known structure and/or motion, gauge fixing is trivial — just arbitrarily choose one point on the given gauge orbit.

- For unknown or varying structure & motion we need to fix a gauge for every possible variation, *i.e.* a local but smooth choice of one point from *each* nearby orbit

— the points form an $(n_{\text{pars}} - n_{\text{gauge}})$-D **local gauge** subspace $\mathcal{G}$, meeting the gauge orbits transversally



Error Minimization Methods for Gauged Problems

- Motions along the gauge are theoretically irrelevant, but cause numerical problems if they become large. So optimization methods for gauged problems must control the gauge motion at each step

- Most numerical gauge control heuristics can be used in two ways :

1 **Fixed gauge** methods try to steer the solution towards a predetermined (but perhaps implicit/adaptive) global gauge

— a good gauge can speed convergence and simplify results

— the need for convergence along the gauge may increase # iterations

2 **Free Gauge** methods let the gauge drift globally, but limit its motion at each step

— less constrained, so a wider choice of numerical methods

Implicit and Parametric Gauges

- $\mathcal{G}$ can be specified in two ways

1 Implicitly by ***local constraints***

$$\boxed{\mathbf{0} \;=\; \mathbf{c}(\mathbf{x}) \;\approx\; \mathbf{c}(0) + \frac{\mathbf{Dc}}{\mathbf{Dx}} \cdot \mathbf{x}}$$

2 By a ***local parametrization*** $\mathbf{u}$ of subspace $\mathcal{G}$:

$$\boxed{\mathbf{x} \;=\; \boldsymbol{g}(\mathbf{u}) \;\approx\; \boldsymbol{g}(\mathbf{0}) + \frac{\mathbf{Dg}}{\mathbf{Du}} \cdot \mathbf{u}}$$

- Compatability :

— $\frac{\mathbf{Dc}}{\mathbf{Dx}}$, $\frac{\mathbf{Dg}}{\mathbf{Du}}$ have full row, column rank

— $\mathbf{c} \circ \boldsymbol{g}(\mathbf{u}) \;=\; \mathbf{0}$ and $\frac{\mathbf{Dc}}{\mathbf{Dx}} \cdot \frac{\mathbf{Dg}}{\mathbf{Dy}} \;=\; \mathbf{0}$ for all $\mathbf{u}$

• Given $\mathcal{G}$, the optimization loops for parameter estimation become

$$\left(\frac{\mathbf{D}g}{\mathbf{Du}}^{\top}\cdot\frac{\mathbf{D}^2 e}{\mathbf{Dx}^2}\cdot\frac{\mathbf{D}g}{\mathbf{Du}}\right)\delta\mathbf{u} \;=\; -\frac{\mathbf{D}g}{\mathbf{Du}}^{\top}\cdot\frac{\mathbf{D}e}{\mathbf{Dx}}$$

$$\begin{pmatrix} \frac{\mathbf{D}^2 e}{\mathbf{Dx}^2} & \frac{\mathbf{Dc}}{\mathbf{Dx}}^{\top} \\ \frac{\mathbf{Dc}}{\mathbf{Dx}} & \mathbf{0} \end{pmatrix}\begin{pmatrix}\delta\mathbf{x} \\ \lambda\end{pmatrix} \;=\; -\begin{pmatrix}\frac{\mathbf{D}e}{\mathbf{Dx}} \\ \mathbf{c}\end{pmatrix}$$

• The corresponding covariance estimates for $\mathbf{x}$ are

$$\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}} \;=\; \frac{\mathbf{D}g}{\mathbf{Du}}\left(\frac{\mathbf{D}g}{\mathbf{Du}}^{\top}\cdot\frac{\mathbf{D}^2 e}{\mathbf{Dx}^2}\cdot\frac{\mathbf{D}g}{\mathbf{Du}}\right)^{-1}\frac{\mathbf{D}g}{\mathbf{Du}}^{\top}$$

$$\begin{pmatrix}\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}} & * \\ * & *\end{pmatrix} \;=\; \begin{pmatrix} \frac{\mathbf{D}^2 e}{\mathbf{Dx}^2} & \frac{\mathbf{Dc}}{\mathbf{Dx}}^{\top} \\ \frac{\mathbf{Dc}}{\mathbf{Dx}} & \mathbf{0} \end{pmatrix}^{-1}$$

WARNING

The dependence of $\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}}$ on $\mathcal{G}$ is often *very strong*

• If the gauge subspaces are strongly tilted *w.r.t.* one another, the same covariance matrix looks completely different in different gauges

— hence, different covariances may not be comparable

• The difference is especially pronounced for camera-centred *vs.* object centred gauges

• It is best to view $\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}}$ as 'nominal' or 'apparent' covariances, rather than absolute ones.

• $\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}}$ is rank deficient — motion along gauge orbits is suppressed by forcing all variations to lie tangent to the local gauge subspace, $\delta\mathbf{x} \in \mathrm{Range}\left(\frac{\mathbf{D}g}{\mathbf{Du}}\right) = \left(\frac{\mathbf{Dc}}{\mathbf{Dx}}\right)^{\perp}$

• $\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}}$ *depends on the chosen gauge* $\mathcal{G}$

— but not on the particular parametrization $\mathbf{x} = \boldsymbol{g}(\mathbf{u})$ or gauge constraints $\mathbf{c}(\mathbf{x}) = \mathbf{0}$ used to represent $\mathcal{G}$.

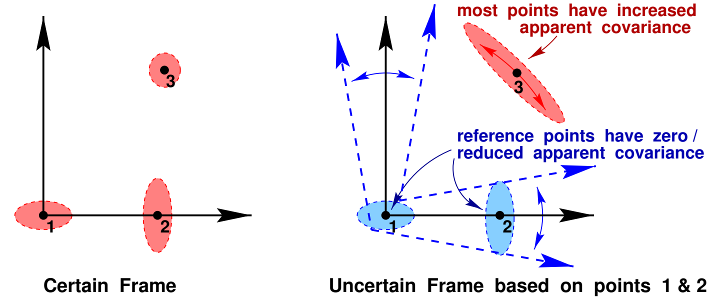
Why is Covariance Gauge Dependent ?

• To fix the gauge, all we can use is scene structure and/or camera motion

• If structure & motion are uncertain, the gauge is also uncertain

• Hence, in SFM, structure, motion & uncertainty estimates are usually expressed in *coordinate frames which are themselves uncertain*

— this confuses the covariance estimates, inflating some, reducing others



• The covariance formalism handles this with deceptive ease.

A more invariant viewpoint

- The gauge $\mathcal{G}$ is imposed from the outside, it is not intrinsic to the problem
- Changes of gauge have no physical significance, so we should not even try to ask for the 'covariance' of gauge dependent quantities
- Instead, view covariance as *an abstract operator defined only on gauge invariant quantities* — *i.e.* those that are constant along gauge orbits
- If we restrict attention to such invariants, any *symmetric two sided pseudoinverse* — $\mathbf{V}$ with $\frac{\mathbf{D}^2 e}{\mathbf{D}\mathbf{x}^2} \cdot \mathbf{V} \cdot \frac{\mathbf{D}^2 e}{\mathbf{D}\mathbf{x}^2} = \frac{\mathbf{D}^2 e}{\mathbf{D}\mathbf{x}^2}$ — can be used as the covariance matrix.
- The space of $\mathbf{V}$'s is exactly the space of local gauge covariances $\mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}}$

Inner Constraint Theorem

- Let $\mathbf{W}$ be a given semi-positive weight matrix, and $\mathbf{N}$ be any $n \times c$ matrix spanning the local gauge directions.

— The gauge $\mathcal{G}$ that minimizes the weighted average covariance $\mathrm{Trace}\left(\mathbf{W} \cdot \mathrm{Cov}\left[\hat{\mathbf{x}}\right]_{\mathcal{G}}\right)$ is given by gauge constraints

$$\boxed{\mathbf{c}(\mathbf{x}) \equiv \mathbf{N}^\top \mathbf{W}\,\mathbf{x} = \mathbf{0}}$$

— $\mathbf{N}$ is easily found from the gauge group action, or as the right null space of $\frac{\mathbf{D}^2 e}{\mathbf{D}\mathbf{x}^2}$.

- **Example:** To minimize the unweighted average structure covariance at a given SFM estimate, constrain the average feature translation, rotation and scale change to be zero.
- Inner constraints can also be approached via Baarda's **S transformations**, which map results between uncertain gauges.

Inner Constraints

- As the nominal covariances depend so much on the chosen gauge, it is useful to ask for gauges in which they are as "well behaved" as possible

— stabler numerical methods, *e.g.* free bundle adjustment

— designing networks, comparing and reporting covariances

- Photogrammetrists call such gauges **inner constraints**

— *e.g.* if we focus on structure covariances, we get traditional object centred coordinates

'Data Snooping'

- Outlier detection by "old-fashioned hypothesis testing"

— developed in 1960's by geodesist W. Baarda, widely used in photogrammetry

— still very relevant, even with robust approaches

- **Internal reliability** = ability to detect (sufficiently large) gross errors using *internal self-consistency* of the measurements.
- **External reliability** = bounds on the maximal expected effect of *undetectable* gross errors.

Redundancy & Residuals

- **Residual fitting error** is the main indicator of internal inconsistency.
- But there are *always* combinations of measurements that produce large model deviations *without* residual errors

— we must ensure that these are rare under some outlier model

- **Redundancy** is the key to reliability

— no parameter combination should depend critically on just a few measurements

— use least squares error propagation to quantify this

3 The true underlying values $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ are unknown, but we have noisy observations

$$\boxed{\underline{\mathbf{y}} \;=\; \overline{\mathbf{y}} + \boldsymbol{\epsilon}}$$

where the noise vector $\boldsymbol{\epsilon}$ has zero mean and known (up to scale) covariance

$$\boxed{\mathrm{Mean}\left[\boldsymbol{\epsilon}\right] \;=\; \mathbf{0} \qquad \mathrm{Cov}\left[\boldsymbol{\epsilon}\right] \;=\; \mathbf{W}}$$

4 We want a good **point estimate** $\hat{\mathbf{x}}$ for $\overline{\mathbf{x}}$ given $\underline{\mathbf{y}}$

'Standard Model' for Parameter Estimation

1 Nonlinear **forward model** uses **parameters** $\mathbf{x}$ to predict **outputs** $\mathbf{y}$

$$\boxed{\mathbf{y} \;=\; \mathbf{f}(\mathbf{x})}$$

— *e.g.* a big vector of predicted image points $\mathbf{y}$ in terms of a big vector of 3D structure and camera parameters $\mathbf{x}$.

2 $\mathbf{x}$ and/or $\mathbf{y}$ may lie in nonlinear manifolds, but we have at least a local linearization

$$\boxed{\Delta\mathbf{y} \;\approx\; \frac{\mathbf{Df}}{\mathbf{Dx}} \cdot \Delta\mathbf{x}}$$

— From now on, we drop the Δ's and let $\mathbf{x}, \mathbf{y}$ denote the linearized local coordinates.

Least Squares Estimation

- **Covariance weighted least squares** gives the Best Linear Unbiased Estimate

— also, for Gaussian noise, the Maximum Likelihood estimate

1 Whiten the noise: decompose $\mathbf{W}^{-1} \;=\; \mathbf{R}^\top\mathbf{R}$ and premultiply by $\mathbf{R}$

$$\boxed{\begin{array}{rcll} \mathbf{R}\,\boldsymbol{\epsilon} & = & \mathbf{R}\left(\frac{\mathbf{Df}}{\mathbf{Dx}}\,\mathbf{x} - \underline{\mathbf{y}}\right) \;=\; \mathbf{A}\,\overline{\mathbf{x}} - \underline{\mathbf{z}} & \\ \mathbf{A} & \equiv & \mathbf{R}\,\frac{\mathbf{Df}}{\mathbf{Dx}} & \text{Normalized Jacobian} \\ \mathbf{z} & \equiv & \mathbf{R}\,\mathbf{y} & \text{Normalized output / observation} \end{array}}$$

- $\mathbf{R}\,\boldsymbol{\epsilon}$ has covariance $\mathbf{I}$, *i.e.* each row is i.i.d
- For simplicity, we use normalized Jacobians and outputs $\mathbf{A}, \mathbf{z}$ from now on.

$\boxed{2}$ Estimate $\overline{\mathbf{x}}$ by minimizing $\|\mathbf{A}\,\hat{\mathbf{x}} - \underline{\mathbf{z}}\|^2$

$$\hat{\mathbf{x}} \equiv \mathbf{A}^\dagger \underline{\mathbf{z}} \quad \text{where} \quad \mathbf{A}^\dagger = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$$
$$\mathrm{Cov}\,[\hat{\mathbf{x}}] = \mathbf{A}^\dagger\,(\mathbf{A}^\dagger)^\top$$

- This is just standard linear least squares
- There are many numerical methods : Normal equations, QR, SVD . . .
- Explicit calculation of $\mathbf{A}^\dagger$ is not usually necessary
- Prior information about $\overline{\mathbf{x}}$ can be encoded as additional 'observations' $\underline{\mathbf{z}}$

Error Propagation

- *Any* perturbation $\delta\underline{\mathbf{z}} \equiv \mathbf{R}\,\delta\underline{\mathbf{y}}$ (noise, outliers . . .) causes perturbations of $\hat{\mathbf{x}}$, $\hat{\mathbf{z}}$, $\hat{\mathbf{r}}$

$$\begin{aligned}\delta\hat{\mathbf{x}} &= \mathbf{A}^\dagger\,\delta\underline{\mathbf{z}}\\ \delta\hat{\mathbf{z}} &= \mathbf{A}\,\mathbf{A}^\dagger\,\delta\underline{\mathbf{z}}\\ \delta\hat{\mathbf{r}} &= \left(\mathbf{I} - \mathbf{A}\,\mathbf{A}^\dagger\right)\,\delta\underline{\mathbf{z}}\end{aligned}$$

- $\delta\underline{\mathbf{z}}$ splits orthogonally into **model** and **residual error** perturbations

$$\delta\underline{\mathbf{z}} = \mathbf{A}\,\delta\hat{\mathbf{x}} + \delta\hat{\mathbf{r}}$$

Residual Error

- The estimated **normalized output** $\hat{\mathbf{z}}$ and **residual error** $\hat{\mathbf{r}}$ are

$$\hat{\mathbf{z}} = \mathbf{A}\,\hat{\mathbf{x}} \qquad \mathrm{Cov}\,[\hat{\mathbf{z}}] = \mathbf{A}\,\mathrm{Cov}\,[\hat{\mathbf{x}}]\;\mathbf{A}^\top = \mathbf{A}\mathbf{A}^\dagger$$
$$\hat{\mathbf{r}} \equiv \underline{\mathbf{z}} - \hat{\mathbf{z}} = \left(\mathbf{I} - \mathbf{A}\,\mathbf{A}^\dagger\right)\underline{\mathbf{z}} \qquad \mathrm{Cov}\,[\hat{\mathbf{r}}] = \left(\mathbf{I} - \mathbf{A}\,\mathbf{A}^\dagger\right)$$

- $\mathbf{A}\,\mathbf{A}^\dagger = \mathbf{A}\,(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ is orthogonal projection onto $\mathrm{Range}\,(\mathbf{A})$
- $\left(\mathbf{I} - \mathbf{A}\,\mathbf{A}^\dagger\right)$ is orthogonal projection onto $\mathrm{Ker}\,(\mathbf{A})$

Redundancy Numbers

- Assuming that all of the error is in a single coordinate $\delta\mathbf{x}_i$, how much of the residual can we "explain"
- The diagonal elements $\rho_i = (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)_{ii}$ are called **redundancy numbers**
- The **total redundancy** $\sum_i \rho_i = \dim\mathbf{y} - \dim\mathbf{x}$ is used in goodness-of-fit estimators like χ^2, AIC
- We want a finer, per-measurement test.
- As above, a fraction ρ_i of an error $\delta\underline{\mathbf{z}}_i$ shows up in the residual $\delta\hat{\mathbf{r}}_i$, so we can estimate

$$\delta\underline{\mathbf{z}}_i \approx \frac{\delta\hat{\mathbf{r}}_i}{\rho_i}$$

— the null (Gaussian i.i.d.) hypothesis suggests that this has distribution roughly χ^2_1

• If ρ_i is small, an error $\delta\underline{\mathbf{z}}_i$ contributes strongly to perturbing the model, but very little to the residual error.

• Any statistical test or robust cost function based on the residual $\hat{r}r_i$ that ignores this reduction of covariance will *accept many outliers*.

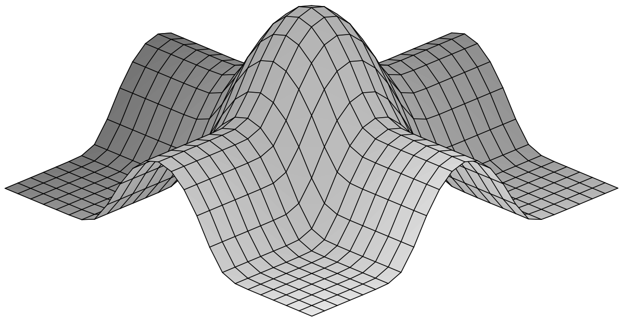
• Given a threshold on $|\delta\underline{\mathbf{z}}_i|$ obtained from $\delta\hat{\mathbf{r}}_i/\rho_i$, we can obtain a bound on the ***maximum expected model perturbation*** caused by an *undetected* error $\delta\underline{\mathbf{z}}_i$

$$\boxed{\|\mathbf{A}\delta\hat{\mathbf{x}}\|^2 \;\approx\; \frac{1-\rho_i}{\rho_i}|\delta\hat{\mathbf{r}}_i|}$$

• We want to control the model error $\delta\hat{\mathbf{x}}$ but it is *unobservable* — we can only measure the residual $\delta\hat{\mathbf{r}}$

• For Gaussian noise $\delta\hat{\mathbf{r}}$ gives no control over $\delta\hat{\mathbf{x}}$ — they are independent.

• But ***correlated outliers are rare*** so for large perturbations $\delta\underline{\mathbf{z}}$ caused by outliers, $\delta\hat{\mathbf{r}}$ *does* give information about $\delta\hat{\mathbf{x}}$



A More Sensitive Outlier Detector

• Suppose there is an outlier in some observation subspace $\mathcal{U}$, how much of the observed residual $\hat{\mathbf{r}}$ could a perturbation $\delta\underline{\mathbf{z}} \in \mathcal{U}$ "explain"

— *e.g.* $\mathcal{U}$ might span the (normalized) observations associated with a single image point, a 3D point in all images, . . .

• Can formulate this as a 'small' least squares problem $\min_{\mathbf{u}\in\mathcal{U}} \left\|\left(\mathbf{I}-\mathbf{A}\mathbf{A}^\dagger\right)\mathbf{u}-\hat{\mathbf{r}}\right\|^2$ with solution

$$\boxed{\mathbf{u} \;=\; \mathbf{U}\left(\left(\mathbf{I}-\mathbf{A}\mathbf{A}^\dagger\right)\mathbf{U}\right)^\dagger\hat{\mathbf{r}}}$$

where $\mathrm{Range}\left(\mathbf{U}\right) \;=\; \mathcal{U}$

• Under the null (Gaussian) hypothesis, $\|\mathbf{u}\|^2$ has distribution $\chi^2_{\dim\mathcal{U}}$

— if $\mathbf{U}$ is orthogonal, $\|\mathbf{u}\|^2 \;=\; \hat{\mathbf{r}}^\top\mathbf{U}\left(\mathbf{U}^\top(\mathbf{I}-\mathbf{A}\mathbf{A}^\dagger)\mathbf{U}\right)^{-1}\mathbf{U}^\top\hat{\mathbf{r}}$

Rules of Thumb for Project Planning

• For multi-image stereo, the 3D accuracy is roughly

$$\boxed{\frac{\text{3D feature error}}{\text{3D object size}} \approx \frac{1}{s\cdot\sqrt{\text{\# views per feature}}}\cdot\frac{\text{feature error in pixels}}{\text{image object size in pixels}}}$$

• **Geometric strength** $s \lesssim 1$ characterizes the camera geometry.

• For cameras subtending half angle θ :

— $s \lesssim \mathcal{O}\left(\sin(\theta)\right)$ for z coordinates,

— $s \lesssim \mathcal{O}\left(\cos(\theta)\right)$ for x, y coordinates.

• Averaging noise by adding more views increases accuracy as $\frac{1}{\sqrt{\text{\# views per feature}}}$

— true even when several views are at the same viewpoint,

— false if accuracy is limited by systematic effects, not image noise.

Maximizing the Geometric Strength

- Use ***strongly convergent*** ($\approx 90^\circ$) geometries with ≥ 3 cameras viewing each feature

  — adding more views beyond a "cone of 4" averages noise but does not significantly increase strength

  — camera distance / focal length has little influence, provided object fills image

  — *c.f.* bas relief increases depth errors for narrow, distant geometries

- Maximize ***overlap of views***, especially for long objects

  — include 'corner' views to span several faces.

- $\approx 90^\circ$ cyclo-rotations greatly improve self-calibration of skew and aspect ratio.