Practical Least-Squares for Computer Graphics

Siggraph Course 11

1. Least Squares with Generalized Errors
   - Weighted Least Squares
   - Total Least Squares
2. Robust Least Squares
   - Motivation
   - Non-Gaussian Distributions
   - Robust error measures
   - Iteratively Reweighted Least Squares
   - Motivation
   - Motivation
   - Least Median of Squares
3. Constrained Least Squares
   - Lagrange Multipliers
   - Non-Negative Least Squares
   - Inequality Constraints

Sermon

Motivation moment

Weighted Least Squares

Extend least squares to account for data with different noise variance per-sample, or missing data

$$\arg\min_{\mathbf{x}} \sum_{i=1}^{n} \frac{\left(\sum_{j=1}^{m} A_{i,j} x_j - b_i\right)^2}{\sigma_i^2}.$$

Weighted Least Squares

$$\arg\min_{\mathbf{x}} \sum_{i=1}^{n} \frac{\left(\sum_{j=1}^{m} A_{i,j}x_j - b_i\right)^2}{\sigma_i^2}.$$

rewrite in matrix terms with $\mathbf{W}$ being a diagonal matrix $\mathbf{W}_{ii} = \frac{1}{\sigma_i}$

$$\Rightarrow \arg\min_{\mathbf{x}} (\mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x}))^T(\mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x})),$$

Weighted Least Squares

rewrite once more

$$\arg\min_{\mathbf{x}} (\mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x}))^T (\mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x})),$$

$$\Rightarrow \arg\min_{\mathbf{x}} (\mathbf{b} - \mathbf{A}\mathbf{x})^T \mathbf{W}^T \mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x}))$$

Rule: $(ab)^T = b^T a^T$

Weighted Least Squares

Big picture:

$$\arg\min_{\mathbf{x}} (\mathbf{b} - \mathbf{A}\mathbf{x})^T \mathbf{W}^T \mathbf{W} (\mathbf{b} - \mathbf{A}\mathbf{x})),$$

This is a "scalar" (a single number), expressing the summed weighted error.

Take the derivative with respect to **x** and set to zero.

The solve for **x**.

Matrix calculus I

- derivative of scalar w.r.t scalar is scalar
- derivative of scalar w.r.t vector is vector
- derivative of scalar w.r.t matrix is matrix

$$\frac{ds}{d\mathbf{x}} = [\frac{ds}{dx_1}, \frac{ds}{dx_2}, \frac{ds}{dx_3}, \cdots]$$

$$\frac{d}{d\mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x}$$

Matrix calculus II

"scalar"	matrix	
$\frac{d}{dx}x^2 \Rightarrow 2x$	$\frac{d}{dx}\mathbf{x}^T\mathbf{x} \Rightarrow 2\mathbf{x}$	
$\frac{d}{dx}ax^2 \Rightarrow 2ax$	$\frac{d}{dx}\mathbf{x}^T\mathbf{A}\mathbf{x} \Rightarrow 2\mathbf{A}\mathbf{x}$	$\mathbf{A}$ symmetric
$\frac{d}{dx}ax \Rightarrow a$	$\frac{d}{dx}\mathbf{a}^T\mathbf{x} \Rightarrow \mathbf{a}$	

Weighted Least Squares

Goal:

$$\arg\min_{\mathbf{x}}(\mathbf{b}-\mathbf{A}\mathbf{x})^T\mathbf{W}^T\mathbf{W}(\mathbf{b}-\mathbf{A}\mathbf{x}))$$

Expand

$$\mathbf{b}^T\mathbf{W}^2(\mathbf{b}-\mathbf{A}\mathbf{x})-\mathbf{x}^T\mathbf{A}^T\mathbf{W}^2(\mathbf{b}-\mathbf{A}\mathbf{x})$$

$$\mathbf{b}^T\mathbf{W}^2\mathbf{b}-\mathbf{b}^T\mathbf{W}^2\mathbf{A}\mathbf{x}-\mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{b}+\mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x}$$

Weighted Least Squares

$$\mathbf{b}^T\mathbf{W}^2\mathbf{b} - \mathbf{b}^T\mathbf{W}^2\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{b} + \mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x}$$

$$\Rightarrow \mathbf{b}^T\mathbf{W}^2\mathbf{b} - 2\mathbf{b}^T\mathbf{W}^2\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x}$$

$\mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{b}$ is a scalar, legal to “transpose” a scalar.

Weighted Least Squares

$$\text{“}c \quad + \quad bx \quad + \quad ax^2\text{”}$$

$$\mathbf{b}^T\mathbf{W}^2\mathbf{b} - 2\mathbf{b}^T\mathbf{W}^2\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x}$$

$$\frac{d}{d\mathbf{x}} = 0 - 2\mathbf{A}^T\mathbf{W}^2\mathbf{b} + 2\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x} = 0$$

Weighted Least Squares

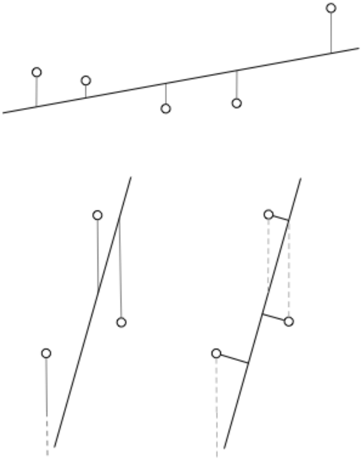
$$\frac{d}{d\mathbf{x}} = 0 - 2\mathbf{A}^T\mathbf{W}^2\mathbf{b} + 2\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x} = 0$$

$$\mathbf{A}^T\mathbf{W}^2\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{W}^2\mathbf{b}$$

$$\mathbf{x} = (\mathbf{A}^T\mathbf{W}^2\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}^2\mathbf{b} = 0$$

Although **A** may not be square, $\mathbf{A}^T\mathbf{W}^2\mathbf{A}$ will be

Total Least Squares



Total Least Squares measures closest error to the (line), rather than in the *y* direction.

Total Least Squares

Unusual: A least squares problem formulation leads to an *eigenvalue* problem rather than a linear system!

Also requires Lagrange multipliers (constrained LS section...).

Total Least Squares: Applications

Surface fitting.

N. Amenta and Y. J. Kil. Defining point-set surfaces, SIGGRAPH 2004.

Reminders

Win a high-end graphics card (HD2900XT) by filling out the course evaluation:

http://www.siggraph.org/courses_evaluation

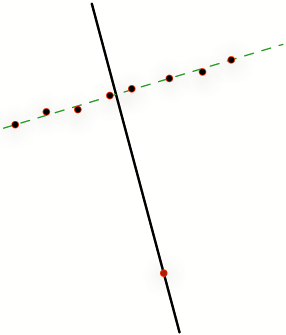
Course web site (download corrected slides after course):

http://www.siggraph.org/courses_evaluation

Robust: Outline

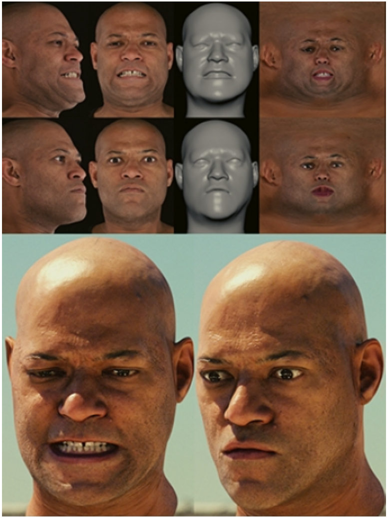
- Motivation
- Redescending error measures
- Iteratively reweighted least squares (IRLS)
- RANSAC
- Least Median of Squares

Robust Least Squares: Motivation



Even a single accidental point (outlier, red point) can destroy an "ordinary" least squares fit.

Robust Least Squares: Applications

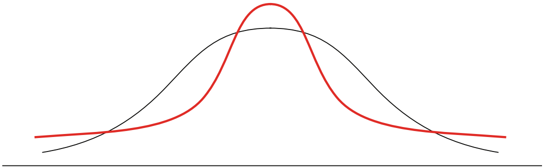


M. Black thesis: introduced robust statistics in optic flow.

Application: optic flow-based face tracking on the *Matrix* sequels

Borshukov et al. Universal Capture - Image-based Facial Animation for “The Matrix Reloaded”

Non-Gaussian Distributions



A high kurtosis density (heavy line) has both more data close to the mean, and more outliers, than a Gaussian distribution (light line).

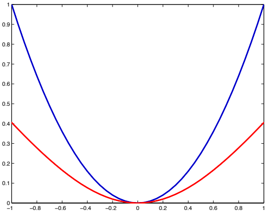
Appropriateness of Gaussian

- Gaussian distribution *is* appropriate when a large number of independent effects are summed (stock market):
- The distribution of a sum is the convolution of the individual distributions. Multiple convolution rapidly converges to Gaussian.

Appropriateness of Gaussian

- Gaussian distribution is not necessarily appropriate when the error is due to a single cause, a few large isolate events, or when the distribution is otherwise simply “non Gaussian”.

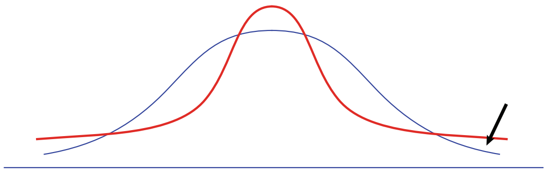
Robust error measures



The *redescending estimator* function $\log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right)$ (red) versus the standard quadratic error $y = x^2$ (blue).

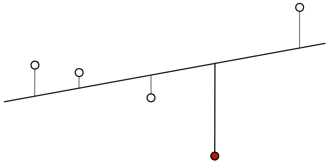
Non-Gaussian Distributions

(repeated figure)



For a high kurtosis error density (heavy line) we want to give less weight to the large errors.

Trimmed Least Squares



A simple approach to robust least squares fitting is to first do an ordinary least squares fit, then identify the k data points with the largest residuals, omit these, perform the fit on the remaining data.

IRLS

Iteratively Reweighted Least Squares generalizes trimmed least squares by raising the error to a power less than 2. No longer "least *squares*".

$$\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_p$$

where $\|\|_p$ is the "L_p" norm, i.e.,

$$\|\mathbf{x}\|_p = \left(\sum \mathbf{x}_k^p\right)^{1/p}$$

The usual least squares minimizes $\|\|_p$ for $p = 2$.

IRLS

The trick is that an error $|e|^p$ can be rewritten as

$$|e|^p = |e|^{p-2}e^2$$

Then, interpret the $|e|^{p-2}$ factor as a weight, and minimize e^2 using weighted least squares.

IRLS: sketch

Problem: e unknown because it depends on x.

iterate:

$$\mathbf{e} = (\mathbf{b} - \mathbf{Ax})$$
$$\mathbf{W} = diag(|\mathbf{e}_i|^{(p-2)/2})$$

solve weighted least squares $\|\mathbf{W}(\mathbf{Ax} - \mathbf{b})\|^2$.

Least Median of Squares (LMedS)

Successful application of Least Median of Squares fitting: The LMedS line (blue) lies close to the model line (black) from which the inliers were generated. The ordinary least squares fit (red line) is influenced by the two outliers.

Least Median of Squares (LMedS)

LMedS finds the fit with the smallest *median* error.

Thus, it can fit data with up to 50% outliers.

Least Median of Squares (LMedS)

Simple algorithm: brute force on a random subset (note slight resemblance to RANSAC).

Repeat M times:

- pick n points at random
- record the median error of the fit on the remaining points

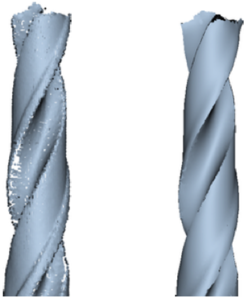
Keep the model with the lowest median error.

Least Median of Squares: SIGGRAPH Applications

Fitting scanned (e.g. Cyberware or LIDAR) data:

Fleishman, Cohen-Or, Silva, Robust Moving Least-Squares Fitting with Sharp Features, Proc. SIGGRAPH 2005 p. 544

Least Median of Squares: SIGGRAPH Applications



Left, Cyberware scan; Right, moving least squares fit initialized with LMedS surface estimate

from: Fleishman, Cohen-Or, Silva, Robust Moving Least-Squares Fitting with Sharp Features, SIGGRAPH 2005

Least Median of Squares: References

- Z. Zhang, Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting, online at `http://www-sop.inria.fr/robotvis/personnel/zzhang/Publis/Tutorial-Estim/Main.html`
- P. Rousseeuw and A. Leroy, Robust Regression and Outlier Detection, Wiley, 1987.

Reminders

Win a high-end graphics card (HD2900XT) by filling out the course evaluation:

http://www.siggraph.org/courses_evaluation

Course web site (download corrected slides after course):

http://www.siggraph.org/courses_evaluation

Lagrange Multipliers

- For *equality* contraints
- Often gives a non-iterative or even closed-form solution

Lagrange Multipliers: Basic Mechanics

1. Express the constraint as $g(\text{variables}) = 0$
2. Add a term $\lambda \cdot g(\text{variables})$ to the original equations.
3. Set derivative with respect to all variables (including λ) = 0, and solve.

Why? (later...)

Lagrange Multipliers: Very Simple Example

Problem: find rectangle of given perimeter with maximum area.

$area = xy$

$perimeter = 2x + 2y$

$\text{constraint} = 2x + 2y = 2$

Lagrange Multipliers: max area given perimeter

$area = xy$

constraint: $2x + 2y = 2 \;\Rightarrow\;$ "$\lambda(2x + 2y - 2)$"

the objective ("Lagrangian")

$$\max_{xy} xy + \lambda(2x + 2y - 2)$$

The constant in the constraint (2 here) usually drops out

Lagrange Multipliers: max area given perimeter

Objective:

$$\max_{xy} xy + \lambda(2x + 2y - 2)$$

$$\frac{d}{dx} = y + 2\lambda = 0 \qquad \Rightarrow \lambda = -\frac{y}{2}$$

$$\frac{d}{dy} = x + 2\lambda = 0$$

$$x + 2 \cdot -\frac{y}{2} = 0 \qquad \Rightarrow x = y$$

Lagrange Multipliers: Somewhat Simple Example

Find a point **p** on a sphere that is closest to a given point **q**.
The constraint that **p** is on a (unit side sphere):

$$\mathbf{p}^T\mathbf{p} = 1$$

Express as $\lambda g(x)$ with $g(x) = 0$:

$$\lambda(\mathbf{p}^T\mathbf{p} - 1)$$

Distance from **p** to **q**: $(\mathbf{p}-\mathbf{q})^T(\mathbf{p}-\mathbf{q})$.
Final cost:

$$\min_{\mathbf{p}}(\mathbf{p}-\mathbf{q})^T(\mathbf{p}-\mathbf{q}) + \lambda(\mathbf{p}^T\mathbf{p} - 1)$$

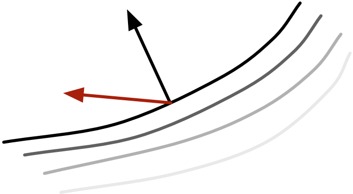
Lagrange Multipliers: Simple Example

Find a point **p** on a sphere that is closest to a given point **q**.
Final cost:

$$\min_{\mathbf{p}} (\mathbf{p}-\mathbf{q})^T(\mathbf{p}-\mathbf{q}) + \lambda(\mathbf{p}^T\mathbf{p}-1)$$

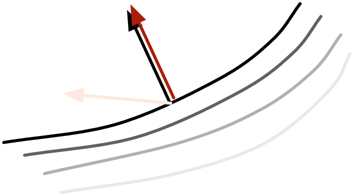
Take derivative with respect to **p**, and λ, and set to zero, obtaining 2 equations. Solve for **p** in the second equation, substitute into the first, result is $\frac{\mathbf{q}}{\|\mathbf{q}\|}$.

Lagrange Multipliers: Intuition



- Minimize $f(x, y)$ subject to $g(x, y) = 0$.
- The solution is at a point where further movement along $g(x, y) = 0$ will not change f. This means there is no component of the gradient of f that is along g, so the gradient of f must be parallel to the gradient of g.

Lagrange Multipliers: Intuition



So

$$\nabla f(x,y) + \lambda \nabla g(x,y) = 0$$

Lagrange Multipliers Example: Inverse Kinematics

$$\mathbf{p} = f(\mathbf{q})$$

p 2d position of end effector, controlled by mouse

q state vector, the n joint angles of a limb.

q has more variables than **p** $\Rightarrow$ *underconstrained*

Lagrange Multipliers Example: Inverse Kinematics

One solution: linearize by taking the derivative with respect to time. (This technique used by Gleicher and Witkin in several papers).

$$\frac{d\mathbf{p}}{dt} = \frac{df}{d\mathbf{q}}\frac{d\mathbf{q}}{dt}$$

and denote $\mathbf{J} \equiv \frac{df}{d\mathbf{q}}$, so

$$\dot{\mathbf{p}} = \mathbf{J}\dot{\mathbf{q}}$$

Lagrange Multipliers Example: Inverse Kinematics

$$\dot{\mathbf{p}} = \mathbf{J}\dot{\mathbf{q}}$$

A linear system, but underconstrained.

Gleicher suggested making the change in state between frames be as small as possible, i.e., minimize

$$\|\dot{\mathbf{q}}\|^2 = \sum_i^n (\dot{\mathbf{q}}_i - 0)^2$$

Lagrange Multipliers Example: Inverse Kinematics

Minimizing $\|\dot{\mathbf{q}}\|^2$ alone would result in no movement.

Instead, minimize it, subject to the constraint that the joint angle change $\mathbf{J}\dot{\mathbf{q}}$ matches the end effector $\dot{\mathbf{p}}$.

This gives the objective

$$\arg\min_{\mathbf{q}} \frac{1}{2}\dot{\mathbf{q}}^T\dot{\mathbf{q}} + \lambda(\dot{\mathbf{p}} - \mathbf{J}\dot{\mathbf{q}})$$

Lagrange Multipliers Example: Inverse Kinematics

Objective

$$\arg\min_{\mathbf{q}} \frac{1}{2}\dot{\mathbf{q}}^T\dot{\mathbf{q}} + \lambda(\dot{\mathbf{p}} - \mathbf{J}\dot{\mathbf{q}})$$

$$\frac{d}{d\dot{\mathbf{q}}} = 0 = \dot{\mathbf{q}} - \mathbf{J}^T\lambda$$

$$\frac{d}{d\lambda} = 0 = \dot{\mathbf{p}} - \mathbf{J}^T\dot{\mathbf{q}}$$

Lagrange Multipliers Example: Inverse Kinematics

$$\frac{d}{d\dot{\mathbf{q}}} = 0 = \dot{\mathbf{q}} - \mathbf{J}^T\lambda$$
$$\frac{d}{d\lambda} = 0 = \dot{\mathbf{p}} - \mathbf{J}^T\dot{\mathbf{q}}$$

Solution: block matrix linear system

$$\begin{bmatrix} \mathbf{J} & 0 \\ \mathbf{I} & -\mathbf{J}^T \end{bmatrix} \begin{bmatrix} \dot{\mathbf{q}} \\ \lambda \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{p}} \\ 0 \end{bmatrix}$$

Lagrange Multipliers Example: Inverse Kinematics

- M. Gleicher, A. Witkin, Differential Manipulation, Graphics Interface 1991
- M. Gleicher, A. Witkin, Through-the-Lens Camera Control, SIGGRAPH 1992

Non-Negative Least Squares: Motivation

$Aw = b$, when A is near singular, can produce very large positive and negative weights.

See **Regularization** section, and also comments in:

- Doug James and Christopher Twigg, Skinning Mesh Animations, SIGGRAPH 2005
- Chuang and Bregler, Performance driven facial animation using blendshape interpolation, Stanford University CS Technical Report.

Non-Negative Least Squares: SIGGRAPH Applications

Skinning:

Doug James and Christopher Twigg, Skinning Mesh Animations, SIGGRAPH 2005

Non-Negative Least Squares: References

- Lawson and Hansen, Solving Least Squares Problems, SIAM, 1995
- Chang, Hyperspectral Imaging, Kluwer 2003 (has good discussion)

Non-Negative Least Squares

- Implementation: nnls.c
- Matlab: lsqnonneg (or lsqlin)

Beyond Non-Negative Least Squares

- Quadratic programming
- Semidefinite Programming

Interlude: basic decomposition/inverse techniques

Though basically all are $O(N^3)$, some are faster than others.

- normal equation: $x = (A^T A)^{-1} A' b$. Fast in the overconstrained case, $A^T A$ is smaller than A.
- QR decomposition. Results in an orthogonal matrix Q and a triangular matrix R. Appropriate for overconstrained (especially if basis for the residual space is needed)
- SVD: slowest, but has many uses, especially in analysis of the problem.

Reminders

Win a high-end graphics card (HD2900XT) by filling out the course evaluation:

http://www.siggraph.org/courses_evaluation

Course web site (download corrected slides after course):

http://www.siggraph.org/courses_evaluation