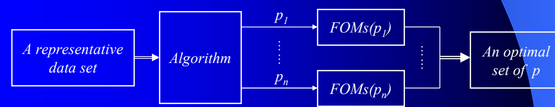
Multi-objective algorithm parameter optimization using multivariate statistics

C.O.S. Sorzano, R. Marabini, G.T. Herman, J.M. Carazo

CSc79000 Presentation
Cecilia Chao Chen

Introduction

- Algorithms
  - *Parameters to determine*
  - *Objective functions to optimize – Figure of Merit (FOM) measures how useful an algorithm in obtain specific information*
- Identify good parameters from simulation experiments



Intro: Pareto Optima

- *N goodness functions $f_1(x), f_2(x), \ldots, f_n(x)$, one variable x*
- *A value a is an Pareto optima iff there is no other value b such that $f_i(a) \leq f_i(b)$ for all $i=1,\ldots,n$ and $f_j(a)<f_j(b)$ for at least one j*
- *An Pareto optimal set is the set of all Pareto optimas*
- *In words, a point is an Pareto optima if it's not lower than another point in all function curves*
- *The variable value at each function's maximum is in Pareto optimal set*
- *Basically a Pareto optima is a trade-off point for all functions, if not the best for all*

Intro: ANOVA studies

- 1-way analysis of variance
  - *An independent variable x and a dependent variable y*
  - *A likelyhood ratio indicates how much the dependent variable is affected by the independent one*
  - *E.g. (teenagers) x=age, y=height; x=SSN, y=height*
  - *Student's t-test : similar purpose; different methods*
- N-way analysis of variance
  - *N independent variables and a dependent variable*
  - *Analyzes how much any or any combination of independent variables influences the dependent variable*

Intro: Principle Components Analysis

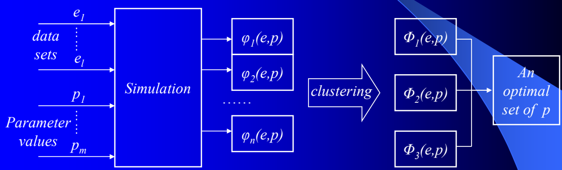
- Principle Components Analysis
  - *Principle components (factors) are the eigenvectors of the covariance matrix of the input data*
  - *The corresponding eigenvalues reflect the amount of variance explained by each principle component*
- Factor Loading
  - *Factor – a linear combination of variables*
  - *Each input data – a linear combination of factors*
  - *Loading – coefficient of each factor*
  - *Introduces a new coordinate system in which the input data vary the most along each axis, thus useful for classification*

Motivations

- Previous work
  - *Only a few FOMs*
  - *Objective function is a formula of parameters*
  - *Standard multi-objective optimization algorithms*
  - *Use Pareto optima as solution*
- Problematic cases
  - *A large number of FOMs to consider*
  - *FOMs randomly distributed due to random noise*
  - *FOM evaluations are time-consuming*
  - *Many parameter values are Pareto-optimal*

Solutions

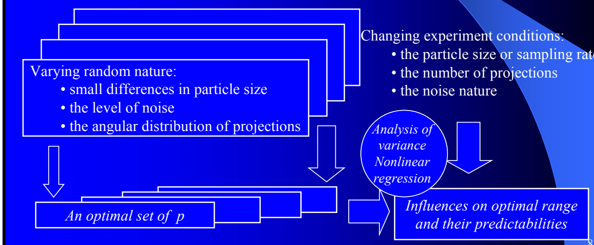
- Statistically based optimization method



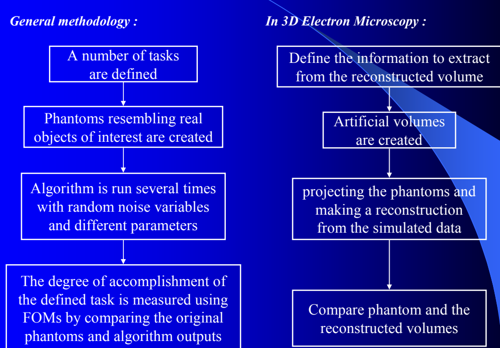
- To illustrate this methodology
  - *3D reconstruction of single particles using ART+blobs*
  - *Decide an optimal range for the relaxation parameter λ*

Solutions

- Investigations in two levels
  - *Intra-experiment optimization (minor changes)*
  - *Inter-experiment analysis (major changes)*



Simulation Procedure



Algorithm for Analysis

- Algebraic Reconstruction Technique with blob
  - *A volume f ~ a linear combination of basis functions b_j*

$$f(\mathbf{r}) \approx \sum_{j=1}^{J} c_j \cdot b_j(\mathbf{r}) = \sum_{j=1}^{J} c_j \cdot b(\mathbf{r}-\mathbf{g}_j)$$

  - *Each measurement y_i ~ corresponding line integral of the basis functions*

$$y_i \approx \sum_{j=1}^{J} l_{i,j} c_j$$

  - *Iteratively update coefficients for basis functions c_i*

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} + \lambda \sum_{i=kM+1}^{(k+1)M} \frac{y_i - \langle \mathbf{l}_i, \mathbf{c}^{(k)} \rangle}{\|\mathbf{l}_i\|^2} \mathbf{l}_i$$

  - *Unspecified parameter: relaxation parameter λ*

Parameter for Analysis

- Parameter to determine

  Relaxation parameter — λ

- Simulation results with reconstruction algorithm ART+blobs

  p_i – value of i^{th} voxel in the phantom

  r_i – value of i^{th} voxel in the reconstruction

  N_v – total number of voxels in the volume

  N_B – total number of voxels in the background

  N_F – total number of voxels in the feature

FOMs for Analysis

- 27 FOMs:
  - *scL2*: l_2-norm of reconstruction error
  - *scL2(B)*: error within background
  - *scL2(F)*: error within feature region
  - *scap*: distance-weighted background error
  - *sccorr*: correlation between two volumes
  - *scinf*: mutual similarity function
  - *scresol*: some resolution measurement ...
  - etc……

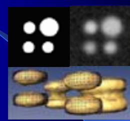
$$scL2 = 1 - \frac{1}{N_V}\sum_{i=1}^{N_V}\left(\frac{p_i - r_i}{2}\right)^2$$

$$scL2(B) = 1 - \frac{1}{N_B}\sum_{i\in B}\left(\frac{p_i - r_i}{2}\right)^2$$

$$scL2(F) = 1 - \frac{1}{N_F}\sum_{i\in F}\left(\frac{p_i - r_i}{2}\right)^2$$

$$scap = 1 - \frac{1}{N_B}\sum_{i\in B} d_i\left(\frac{p_i - r_i}{2}\right)^2$$

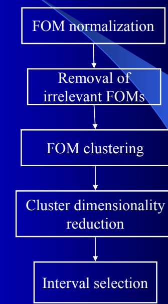
Experimental Settings



- Reconstruction settings
  - *Phantom size: 65 × 65 × 65*
  - *Projection image number: 2010*
  - *Noise: Gaussian additive noise*
- Minor changes
  - Cylinder size: *vertical distance 3~6 voxels; radius 5~10 voxels; height 3~5 voxels*
  - Projection direction: *evenly and randomly distributed along all possible directions*
  - Gaussian noise: *standard deviation 9*
- Simulations
  - *10 simulations for each λ from 0.015 to 0.15 in steps of 0.015; 27 FOMs → 100 × 28 table, each row: λ and 27 FOM values*

Intra-Experiment Optimization

- Optimization steps
  - *Normalize FOMs to comparable values*
  - *Remove FOMs insensitive to changes of parameter value*
  - *Cluster FOMs with similar dependency on parameters*
  - *Obtain a single representative for each cluster*
  - *Optimal region selection based on cluster representatives*

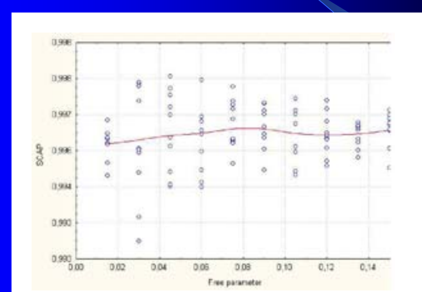


Intra-Experiment Optimization

- FOM normalization
  - *Normalize to zero-mean and unit variance*
  - *Applied independently to each column of FOM values*
  - *An outlier rejection might be needed before normalization*
- Removal of irrelevant FOMs
  - *1-way ANOVA studies*
  - *Independent variable λ; dependent variable FOM*
  - *Remove FOMs with insignificant dependence on λ*
  - *4 FOMs show dependence not significant than 90%*

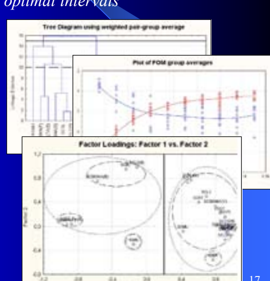
Intra-Experiment Optimization

- Removal of irrelevant FOMs



Intra-Experiment Optimization

- FOM clustering
  - *Identify groups of FOMs with similar behavior with respect to the parameter*
  - *Different groups may point to different optimal intervals*
  - *Use a hierarchical classifier for clustering*
  - *Grouped into two main classes: one decreases with λ, one increases with λ*
  - *One group reflects the brightness while the other reflects the noise level in reconstruction.*
  - *Use Principal Components Analysis for validation: first two factors accumulate more than 87% of the total variance*



Intra-Experiment Optimization

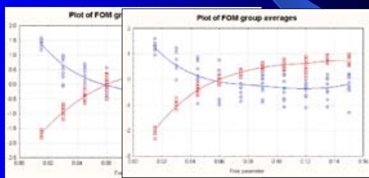
- Cluster dimensionality reduction
  - *FOMs in one group show similar trend*
  - *Reduce all FOMs in one cluster to a single representative*
  - *Principal Components Analysis on each FOM group: the first three components account for group variance 98% and 94% in two groups respectively*
  - *Variance-weighted combine principle components*

$$FOM_G = \Sigma^{P}_{i=1} w_i pc_i$$

(w_i percent of total variance is explained by principle component pc_i)

Intra-Experiment Optimization

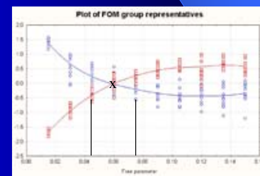
- Cluster dimensionality reduction
  - *Number of objective functions reduced to number of tendencies*



  - *Similar to group average with slightly smaller variance for each λ*

Intra-Experiment Optimization

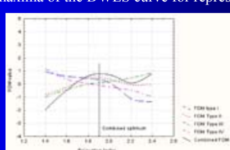
- Interval selection
  - Select the parameter value achieving a compromise among reduced objective functions
    - *Multi-objective optimization searching for the Pareto-optimal solution show all values are valid in ART*
    - *Find a **Middling** point reasonable for all FOM groups*
  - Select the middling interval
    - *Contains the middling point*
    - *In each group, find intervals where representative FOM not significantly different from that at the middling point (Student's t-test)*
    - *Intersection of all the group intervals is the middling interval*



(0.045, 0.075)

Intra-Experiment Optimization

- Interval selection
  - Alternative middling point choices
    - *For Iterative Data Refinement with four FOM groups*
      - Average the FOM group representatives
      - select maxima of the DWLS curve for representatives average



    - *Choose the free parameter that minimizes the variance of the DWLS fitted trends (For ART, chosen point variance=0)*

Inter-Experiment Analysis

- Major changes may affect the optimal interval
  - *Data collection geometry*
  - *The particle size or sampling rate*
  - *The number of projection images*
  - *Noise nature*
- Two questions
  - *Which major changes have an important influence?*
  - *Is the dependency predictable?*
- Experimental settings
  - *Volume size and projection number*

    (65^3, 773) (65^3, 2010) (65^3, 3246)

    (97^3, 1153) (97^3, 2999) (97^3, 4844)

    (129^3, 1533)(129^3, 3988)(129^3, 6443)
  - *Two kinds of noises (SNR=0.33 in both cases)*
    - Unfiltered white Gaussian noise – similar to low defocusing image
    - White Gaussian noise low pass filtered at f=0.2 (f_{max} normalized to 0.5) – similar to strongly defocused

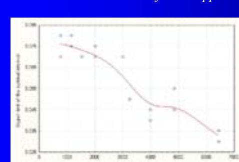
Inter-Experiment Analysis

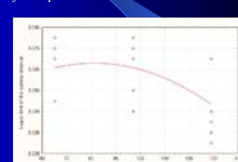
Size (voxels)	Number of projections	Noise	λ_{min}	λ_{max}
65^3	773	unfiltered	0.060	0.080
65^3	2010	unfiltered	0.050	0.070
65^3	3246	unfiltered	0.030	0.050
97^3	1153	unfiltered	0.060	0.080
97^3	2999	unfiltered	0.050	0.070
97^3	4844	unfiltered	0.035	0.055
129^3	1533	unfiltered	0.040	0.020
129^3	3988	unfiltered	0.025	0.045
129^3	6443	unfiltered	0.015	0.030
65^3	773	filtered	0.050	0.070
65^3	2010	filtered	0.055	0.075
65^3	3246	filtered	0.030	0.050
97^3	1153	filtered	0.055	0.075
97^3	2999	filtered	0.050	0.070
97^3	4844	filtered	0.025	0.045
129^3	1533	filtered	0.050	0.070
129^3	3988	filtered	0.020	0.040
129^3	6443	filtered	0.015	0.035

- Which changes are affective?
  - *N-way ANOVA study*
  - *Only noise nature is irrelevant*
  - *The number of projects explains twice more variance than the volume size*
  - *No significant association found among three variables*
- Is the dependency predictable?
  - *Nonlinear Regression Analysis*
  - *Establish a model for the upper limit of the optimal interval*

Inter-Experiment Analysis

- Is the dependency predictable?
  - *Nonlinear Regression Analysis*
  - *Establish a model for the upper limit of the optimal interval*





Model	Formula	R^2
Constant	0.074	0
$f_1(N)$	$0.0842 - 8\cdot10^{-?}N^{?}$ [illegible]	0.82
$f_2(N)$	$f_1(N) - 0.391\cdot10^{-12}(N-1021)(N-2697)(N-4643)(N-6501)$	0.85
$f_3(N,S)$	$f_1(N) + 5.83\cdot10^{-3} - 6\cdot10^{-6}\lvert S-95.76\rvert^{2.12}$	0.89
$f_4(N,S)$	$f_2(N) + 5.44\cdot10^{-3} - 0.8\cdot10^{-6}\lvert S-94.65\rvert^{2.44}$	0.92

$\lambda_{max} = 0.09 - 10^{-6}(8N - 6(S-96)^2)$

$\lambda_{min} = \lambda_{max} - 0.02$

[only valid for particular conditions]

Conclusions

- Multivariate statistics intra-experiment analysis
  - *Recognize different FOM tendencies, decrease complexity of FOM optimization problem, and find a trade-off among all FOMs*
  - *Feature: recognition of statistical associations among variables, to characteristic main problems by different FOMs*
- Inter-experiment analysis
  - *Identify major changes that affect the parameters effectively*
  - *Use nonlinear regression to establish a model of dependency*
- Optimization experiments ART+blobs
  - *Reconstructions using relaxation factor λ within the optimal interval show the best visual appearance, as well as achieved an FOM trade-off*
  - *Inter-experiment analysis shows the filtered or unfiltered nature of the additive noise doesn't affect the parameter optimal interval – indicating no need to distinguish images of different defocusing as far as the concern is selecting λ*
  - *The most important factor determining λ is the number of projection images – more projection means more volume updates and a smaller iterative step needed to reach a certain optimal reconstruction*
  - *the other important factor is the size of volume – increasing volume side length increases quadratically the number of equations specified by the projections, thus lessen the iterative step*

Appendix

- T-Test
  - To test whether variation between two groups is significant

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

- ANOVA
  - To test whether variation within more than two groups is more than random variation

$$SS(Total) = SS(Between) + SS(Within)$$

where $SS(Total) = \sum\sum (X_{ij} - \bar{X})^2$

$SS(Between) = \sum (\bar{X}_j - \bar{X})^2$

$SS(Within) = \sum\sum (X_{ij} - \bar{X}_j)^2$

where $\bar{X}$ is the mean of the all observations and $\bar{X}_j$ is the mean of the group j.

$$MS(Within) = \frac{SS(Within)}{n-k} \qquad MS(Between) = \frac{SS(Between)}{k-1}$$

The mean sum of squares are comparable and the test value

$$\hat{F} = \frac{MS(Between)}{MS(Within)} \sim F(k-1, n-k)$$

is large, if the variation between the groupmeans is larger than the variation in the groups. The significance of the test $p = P(F > \hat{F})$ tells the probability of the H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$.