# Optimizing Expectation

Lingyu Chen
University Oral Examination
Department of Statistics
Stanford University
June 6, 2000

Adviser: Professor Art B. Owen



## Problem

- Seek

$$\text{arg min}_x E_y(f(x, y))$$

  over $x \in \mathbb{R}^p$, where $y \in \mathbb{R}^q$.
- $x$ represents some parameters we can control.
- But uncertain about $y$.



## Examples

| **Domain** | **Choose $x$** | **Random $y$** |
|---|---|---|
| Finance | portfolio | future price |
| Nonlin. exptl. design | design matrix | sampling noise |
| Communication theory | equalizer | background noise |



## Special Cases

- EM
  - Tractable if function in closed form
- Stochastic linear programming (SLP)
  - Linear approximations

## A Naive Solution

- Find

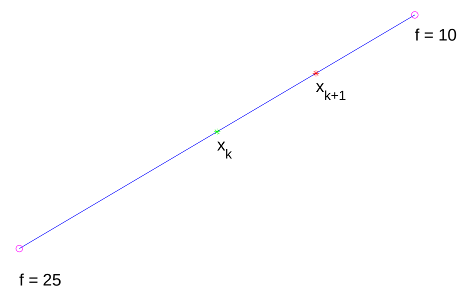
$$\text{arg min}_x \frac{1}{N} \sum_{i=1}^{N} f(x, Y_i)$$

where $Y_i$ is an integration rule.

- But expensive when $N$ is large.
- Slow even for one iteration.
- *Idea:* use smaller $N$ + more iterations



## Extreme Case: N = 2




- Random direction in $\mathbb{R}^p$
- Move towards down hill
- Random Direction Stochastic Approximation(RDSA)



## Robbins-Monro Procedure (1951)

- Seek the root $x^* \in \mathbb{R}$ of

$$F(x) = 0$$

where $F(x) = E[f(x, y)]$ is increasing and unknown.

- Robbins-Monro:

$$x_{n+1} = x_n - a_n f(x_n, y_n)$$

where $a_n > 0$.

- Under general conditions, the convergence rate is $n^{-1/2}$.



## Kiefer-Wolfowitz Procedure (1952)

- Seek $x^* \in \mathbb{R}$,

$$x^* = \text{arg min}_x F(x) = \text{arg min}_x E[f(x, y)]$$

- Kiefer-Wolfowitz:

$$x_{n+1} = x_n - a_n \frac{f(x_n + c_n, y_n^+) - f(x_n - c_n, y_n^-)}{2c_n}$$

where $a_n > 0, c_n > 0, y_n^{\pm}$ random.

- Need
  - $c_n$: search distance
  - $d_n$: move distance

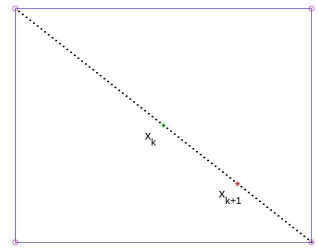
## General Stochastic Approximation

- Now $x \in \mathbb{R}^p$.
- Recall
  - $c_n$ = search distance
  - $d_n$ = moving distance
- More specifically
  - $d_n = a_n \hat{g}_n$
  - $\hat{g}$ = directional derivative in random direction
- Kushner & Yin(1997) showed:
  - $x_n \to x^*$
  - Asymptotic normality
  - error $\sim n^{-1/3}$

  Assuming:
  - $a_n \to 0, \quad \sum a_n = \infty$
  - $c_n \to 0, \quad \sum (\frac{a_n}{c_n})^2 < \infty$
- Typically, $\quad a_n = \frac{a}{n^\alpha}, \quad c_n = \frac{c}{n^\gamma}$.



## Simultaneous Perturbation




- SPSA(Spall 1992): random directions on the vertices of the unit cube
- Only use $2^{p-1}$ choices of directions.
- Controversial efficiency

| Source | Conclusion | Investigation |
|---|---|---|
| Spall | Better | Empirical |
| Kushner & Yin | No Better | Theory |



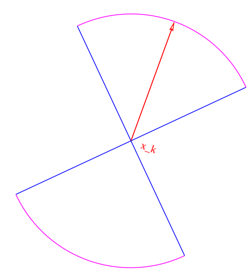
## Statistical Ideas

Iterative process generates data. Use them to:

- Select meta parameters, e.g. $a$ and $c$
- Detect Convergence
- Adapt to important directions



## Momentum




- Identify the locally important directions.
- Locally adapt the search direction
- Change the search region with dimension.
- Search in central $20\%$ of the surface.
- Note in $\mathbb{R}^{100}$, focus on central $20\%$ of the surface when angle $\leq 83°$.
- Move in a favorable direction.

## Noise Coupling

- The estimate of directional derivative along direction $d_n$:

$$\frac{1}{2c_n}[f(x_n + c_n d_n, y_n^+) - f(x_n - c_n d_n, y_n^-)]$$

- We can set $y_n^+ = y_n^-$ in simulation optimization.
- Eliminate the $1/c_n$-dependent term in the effective noise.
- Reduce the variance of estimate of directional derivative.
- Improve rate from $n^{-1/3}$ to $n^{-1/2}$ for related methods.



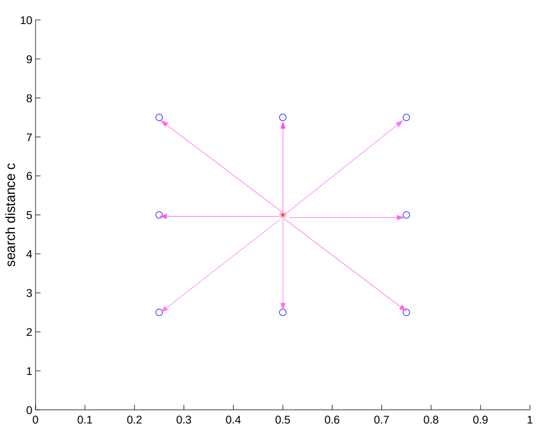
## Meta Parameters

- Search: $c_n = \frac{c}{n^\gamma}$
- Move: $a_n = \frac{a}{n^\alpha}$
- But $a$ and $c$ unknown.
- Start with:
  - $c$: standard deviation of noise
  - $a$: problem specific change magnitude



## Pattern Search

- Pattern search: periodically try several $(a, c)$ pairs.




- Choose apparent winner.
- Spend up to $10\%$ of total function evaluations this way.



## Detection of Convergence

- A function $f$ is antitonic if $f(x_1) \geq f(x_2) \geq \cdots \geq f(x_k)$ with respect to an ordering $x_1 \preceq x_2 \preceq \cdots \preceq x_k$.
- Fit a sample antitonic regression function to the observed values.
- Antitonic regression: the pool-adjacent-violators algorithm(PAVA).
- Test the hypothesis that the function values are antitonic.
- Investigate statistical approach of detecting whether the iterates have reached the neighborhood of the solution.

## Portfolio Example

- Constant rebalanced portfolio
- Allocate money among $p$ assets.
- Put fraction $x_i \geq 0$ in asset $i$, $\sum_1^p x_i = 1$.
- $y_{it}$: return to asset $i$ in period $t$
- Asset returns follow a lognormal distribution.



## Log-optimum Investment

- Log-optimal portfolio(Breiman 1960):

$$\max_x E\Big(\log \prod_{t=1}^{T} \sum_{i=1}^{p} x_i y_{it}\Big)$$

- Assume returns are iid through time.
- Cover's (1984) algorithm:

$$x_i^{n+1} = E\Big(\frac{x_i^n y_i}{\sum_{i=1}^p x_i^n y_i}\Big)$$

- Need simulation in order to handle
  - Loads on purchases
  - Taxes on sales
  - Non-normal distribution

    heavy tails, stochastic volatility
  - More complicated strategies



## Simulation

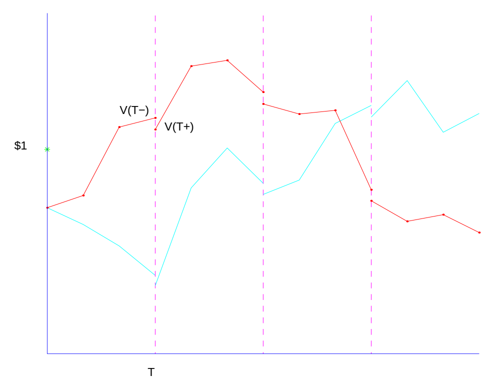
- Start with $1.

  $x_j$ in asset $j$, $\sum_{j=1}^p x_j = 1$
- Sample returns.

  $x_j^{'}$ in asset $j$, $\sum_{j=1}^p x_j^{'} = 1$, $x_j^{'} = \frac{x_j y_j}{\sum_k x_k y_k}$
- Rebalance to $x_j$ after each period.



## Rebalancing with Loads



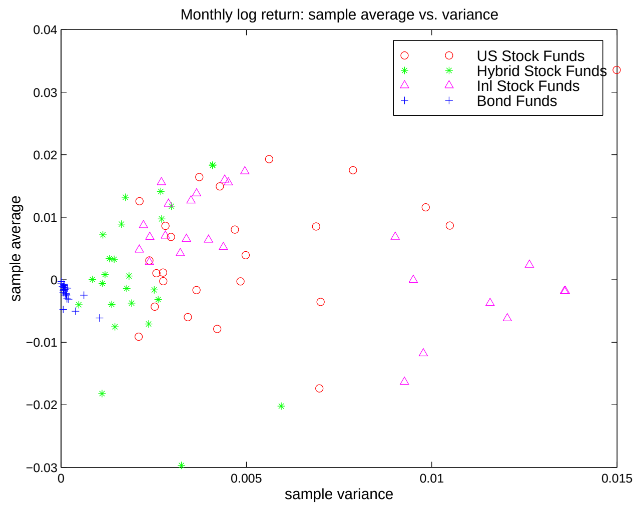

- Time $T-$ have $V(T-)x_j^{'}$ in asset $j$

  $T+$ have $V(T+)x_j$ in asset $j$
- Load $c_j$ on the purchase of asset $j$
- Optimally $V(T+) = (1-\alpha)V(T-)$ with

$$\alpha = \sum_{(1-\alpha)x_j > x_j^{'}} c_j((1-\alpha)x_j - x_j^{'})$$

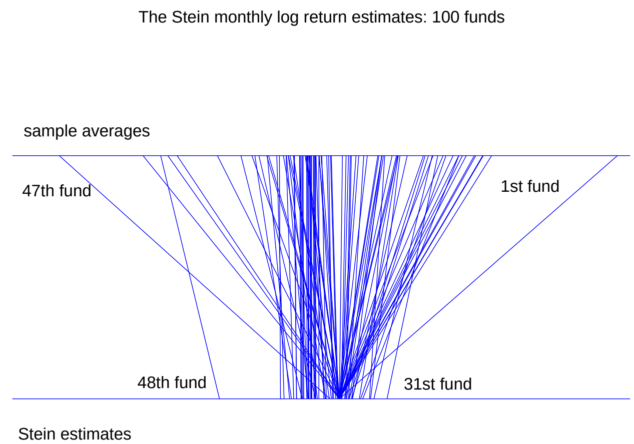
## Historical Returns

- Use historical returns of $100$ mutual funds(Bloomberg) from Jan. 1997 to Jan. 2000.





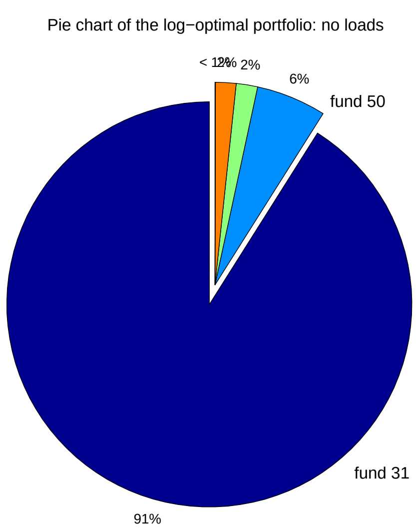

## Structure Portfolio






## Log-optimum: No Loads

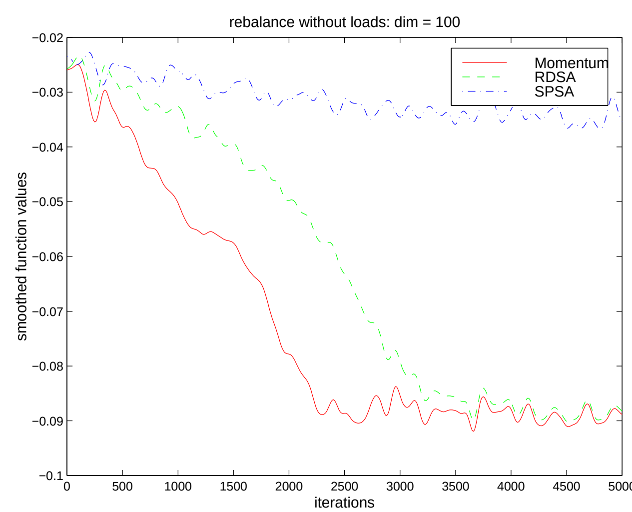
- The log-optimum investment over $1$ year
- Quarterly rebalancing
- Invest $91\%$ of money in fund $31$.





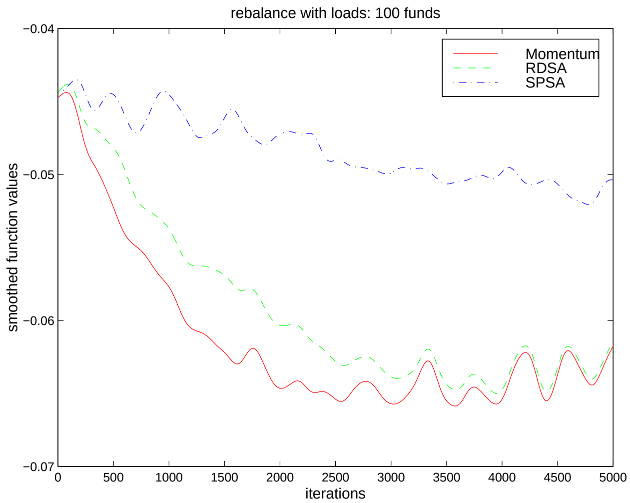

## Comparison of SA algorithms (No Loads)

- Minimize the minus expected log return.



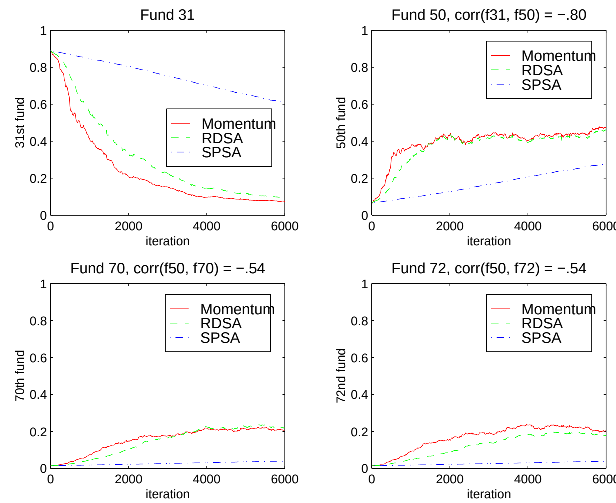
## Comparison of SA algorithms (with Loads)

- Charge $3\%$ loads on some funds.





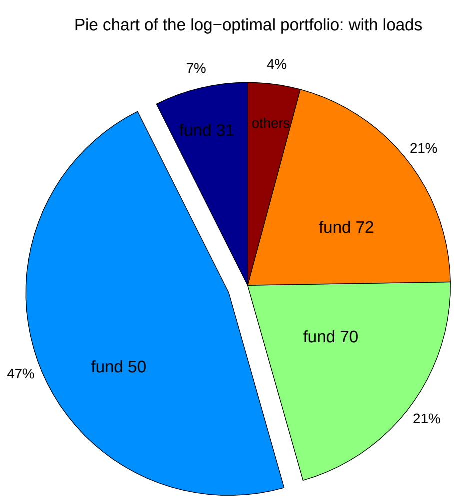

## Log-optimal with Loads: Trajectories






## Log-optimum: with Loads

- Prefer no-load funds.
- Invest in negatively correlated funds.






## Bayesian Experimental Design

- Priors on the parameters
- Maximize an expected utility function
- Linear design:
  - Alphabetical optimality
  - An upper bound on the # of support points exists.
- Nonlinear design:
  - Expected utilities have no closed form.
  - No bound on the # of support points

## Logistic Regression

- Binary responses
- The probability of success:

$$\frac{\exp(\beta(x-\mu))}{1+\exp(\beta(x-\mu))}$$

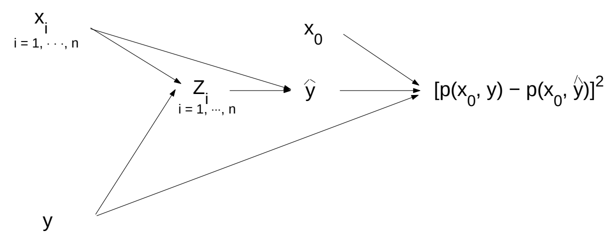
- Here $y = (\mu, \beta)^T$
  $x =$ support points
- Uniform and independent priors for $\mu$ and $\beta$.
- Choose optimal $x$ for different criteria.



## Squared Prediction Error

- Minimize expected squared prediction error:

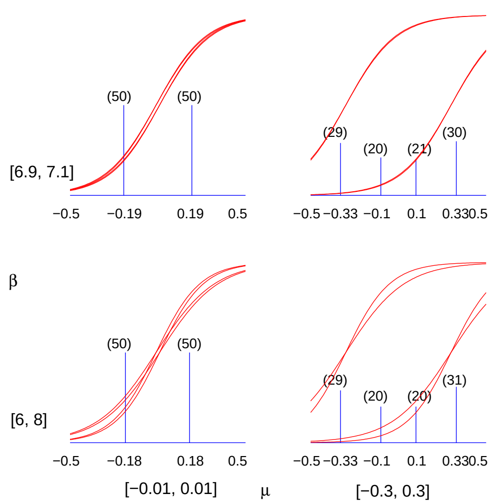
$$\min_{x_i, 1\le i\le n} E_y E_{Z_1,\cdots,Z_n|x,y} E_{x_0}(p(x_0,\hat{y})-p(x_0,y))^2$$






## Prediction: Symmetric Test Region

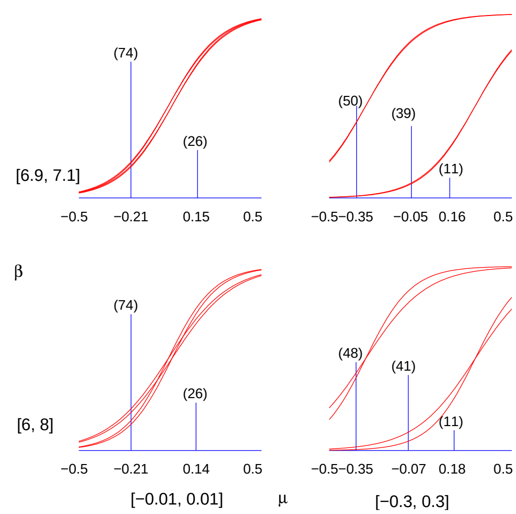
- Test region: $[-0.5, 0.5]$






## Prediction: Asymmetric Test Region

- Test region: $[-0.5, 0]$



## Summary

- Use data to improve the performance of SA.
- The momentum algorithm appears to be the fastest, especially in large dimensional problems.
- Some practical issues, e.g, the selection of algorithm parameters, are critical to the finite-sample performance.



## Future Work

- Asymptotic theory for the momentum algorithm
- Constrained optimization
- Second order stochastic approximation
- Capital gains taxes in portfolio selection
- Nonlinear experimental design in high dimensional space