Multidirectional subspace expansion for single- and multi-parameter Tikhonov regularization

SIAM LA15

Ian Zwaan

Department of Mathematics and Computer Science
Eindhoven University of Technology

October 27, 2015

Introduction

What?

Consider a (large scale) least squares problem

$$A\boldsymbol{x} \approx \boldsymbol{b}, \quad \boldsymbol{b} = A\boldsymbol{x}_\star + \boldsymbol{n}, \quad \|\boldsymbol{n}\| = \epsilon,$$

where A is sparse and ill-conditioned $\rightsquigarrow$ need regularization.

We will investigate general form Tikhonov regularization

$$\arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2,$$

and multi-parameter Tikhonov regularization

$$\arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \sum_{i=1}^{\ell} \mu^i \|L^i \boldsymbol{x}\|^2.$$

Introduction

Why?

Usually use general form or multi-parameter for better solutions.

- $\mathcal{N}(L)$ is not penalized
- L can dampen unwanted properties
- L^i when additional prior information is available

General form can be transformed to standard form:

$$\arg\min_{\boldsymbol{y}} \|A(I - (A(I - L^{\dagger}L))^{\dagger}A)L^{\dagger}\boldsymbol{y} - \boldsymbol{b}\|^2 + \mu\|\boldsymbol{y}\|^2,$$

$$\boldsymbol{y} = L\boldsymbol{x},$$

but can be cumbersome to compute and deal with.

Impossible for multi-parameter.

Review

Standard form

Standard form Tikhonov regularization ($\mu > 0$)

$$\arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|\boldsymbol{x}\|^2$$

Create orthogonal bases U_{k+1} and V_k with Golub–Kahan–Lanczos bidiagonalization:

$$A^* U_k = V_k B_k^*$$
$$AV_k = U_{k+1}\bar{B}_k$$

$\bar{B}_k$ is lower-bidiagonal and B_k is the upper square part of $\bar{B}_k$.

Review

Standard form

New problem becomes

$$\begin{aligned}
&\min_{\boldsymbol{x}\in\mathcal{V}_k} \|A\boldsymbol{x}-\boldsymbol{b}\|^2+\lambda\|\boldsymbol{x}\|^2 \\
&=\min_{\boldsymbol{c}} \|AV_k\boldsymbol{c}-\boldsymbol{b}\|^2+\mu\|V_k\boldsymbol{c}\|^2 \\
&=\min_{\boldsymbol{c}} \|\bar{B}_k\boldsymbol{c}-\beta\boldsymbol{e}_1\|^2+\mu\|\boldsymbol{c}\|^2.
\end{aligned}$$

Now $\boldsymbol{x}_k = V_k\boldsymbol{c} \in \mathcal{V}_k$, where

$$\begin{aligned}
\mathcal{V}_k &= \mathrm{span}(V_k) \\
&= \mathcal{K}_k(A^*A, A^*\boldsymbol{b}) \\
&= \mathrm{span}\{A^*\boldsymbol{b}, (A^*A)A^*\boldsymbol{b}, \dots, (A^*A)^2A^*\boldsymbol{b}, (A^*A)^{k-1}A^*\boldsymbol{b}\}.
\end{aligned}$$

Review

General form

Consider again general form Tikhonov regularization

$$\arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2$$

One possibility is to generate V_k as before and compute

$$LV_k = W_k K_k$$

and

$$\begin{aligned}
&\min_{\boldsymbol{x} \in \mathcal{V}_k} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2 \\
&= \min_{\boldsymbol{c}} \|AV_k\boldsymbol{c} - \boldsymbol{b}\|^2 + \mu\|LV_k\boldsymbol{c}\|^2 \\
&= \min_{\boldsymbol{c}} \|\bar{B}_k\boldsymbol{c} - \beta\boldsymbol{e}_1\|^2 + \mu\|K_k\boldsymbol{c}\|^2.
\end{aligned}$$

Review

Another approach

Lampe, Reichel, and Voss (2012) compute

$$AV_k = U_{k+1}\bar{H}_k \qquad LV_k = W_k K_k$$

as before, but expand with

$$A^*b - (A^*A + \mu_k L^*L)V_k\mathbf{c}.$$

Has the following nice properties:

- orthogonal to V_k in exact arithmetic
- gradient of objective function in $\mathbf{x}_k$
- residual of the normal equations

Easy to extend to multi-parameter regularization

$$A^*b - (A^*A + \sum_{i=1}^{\ell} \mu_k^i L^{i*}L^i)V_k\mathbf{c}.$$

Only one new basis vector per iteration!

Multidirectional subspace expansion.

Multidirectional expansion

Idea

Effectively expand with

$$(A^*A + \sum_{i=1}^{\ell} \mu_k^i L^{i*} L^i) V_k \mathbf{c}.$$

Q: Optimal? Or can we find a "better" linear combination of

$$A^*A\mathbf{x}_k, \quad L^{1*}L^1\mathbf{x}_k, \quad \ldots, \quad L^{\ell*}L^{\ell}\mathbf{x}_k.$$

A: Not (reliably) without extra MVs.

With extra MVs: expand with each term.

Downside: $\ell + 1$ new basis vectors per iteration!
Solution: keep only the "best" and remove ℓ vectors per it.
N.B.: ℓ is often small, e.g., $\ell \leq 3$.

Multidirectional expansion

Step-by-step

In iteration $k+1$

- Expand basis with $\ell+1$ new vectors and obtain $V_{k+\ell+1}$.
- Select regularization parameters $\boldsymbol{\mu}_{k+1}$ and compute $\mathbf{c}_{k+\ell+1}$.
- Compute orthonormal matrix Z such that $Z\mathbf{c}_{k+1:k+\ell+1} = \xi\mathbf{e}_1$.
- Observe that

$$V_{k+\ell+1}\mathbf{c}_{k+\ell+1} = V_k\mathbf{c}_{1:k} + V_{k+1:k+\ell+1}Z^*(\xi\mathbf{e}_1).$$

- Multiply the last $\ell+1$ columns of $V_{k+\ell+1}$ by Z^*.
- Multiply the last $\ell+1$ columns of $\bar{H}_{k+\ell+1}$ by Z^*.
- Multiply the last $\ell+1$ columns of $K^i_{k+\ell+1}$ by Z^*.

Multidirectional expansion

Step-by-step

Need to make $\bar{H}_{k+\ell+1}$ UH and $K^i_{k+\ell+1}$ UT.

- Compute P such that $P\bar{H}_{k+2:k+\ell+2,k+1:k+\ell+1}$ is UT.
- Compute Q such that $QK^i_{k+1:k+\ell+1,k+1:k+\ell+1}$ is UT.
- Apply P and Q to the bottom rows of $\bar{H}_{k+\ell+1}$ and $K^i_{k+\ell+1}$.
- Multiply the last $\ell+1$ columns of $U_{k+\ell+2}$ by P^*.
- Multiply the last $\ell+1$ columns of $W^i_{k+\ell+1}$ by Q^*.
- We now have a similar decomposition as at the start

$$AV_{k+\ell+1} = U_{k+\ell+2}\bar{H}_{k+\ell+1}, \quad L^iV_{k+\ell+1} = W^i_{k+\ell+1}K^i_{k+\ell+1},$$

but with our approximation in $\mathrm{span}(V_{k+1})$.

Multidirectional expansion

Example

Expand V_1 with $A^*A\mathbf{v}_1$ and $L^*L\mathbf{v}_1$. Compute

$$AV_{1+2} = U_{2+2}\bar{H}_{1+2}, \quad LV_{1+2} = W_{1+2}K_{1+2}$$

Select μ_2 and compute $\mathbf{c}_{1+2}$, Z, P, and Q.

$$\xrightarrow{\bar{H}_{1+2}} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\bar{H}_{2:3}Z^*} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{P\bar{H}_{3:4,2:3}Z^*} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

$$\xrightarrow{K_{1+2}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{K_{2:3}Z^*} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{QK_{2:3,2:3}Z^*} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

Now truncate.

Parameter selection.

Parameter selection
Review

Methods for one parameter: L-Curve, GCV, ..., discrepancy.

For discrepancy solve

$$\phi_k(\mu) = \|A\boldsymbol{x}_k(\mu) - \boldsymbol{b}\|^2 = (\eta\epsilon)^2 \qquad (\eta > 1),$$

where

$$\boldsymbol{x}_k(\mu) = \arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2.$$

Easy with SVD or GSVD.

Problem for multiple parameters...

Parameter selection
Review

Discrepancy for multi-parameter

- Brezinski, Kilmer, and Miller (2003)
- Lu, Pereverzev, Shao, and Taunenhahn (2011)
- Gazzola and Novati (2013)

Parameter selection

Ideas

Consider

$$\arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu \left(\sum_{i=1}^{\ell} \omega^i \|L^i \boldsymbol{x}\|^2 \right).$$

Choose μ s.t. the discrepancy principle is satisfied.

Choose ω^i to "undo" scaling and more?

Let μ^i s.t. $\phi^i(\mu^i) = (\eta\epsilon)^2$ and

$$\boldsymbol{x}^i(\mu) = \arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2$$

Take

$$\omega^i = \frac{1}{\|D\boldsymbol{x}^i(\mu^i)\|}$$

Parameter selection

Continued

What happens if we scale any of the terms?

$$\arg\min_{\tilde{\boldsymbol{x}}} \|\alpha A\tilde{\boldsymbol{x}} - \beta \boldsymbol{b}\|^2 + \sum_{i=1}^{\ell} \tilde{\mu}^i \|\lambda^i L^i \tilde{\boldsymbol{x}}\|^2$$

The noisy component of $\beta \boldsymbol{b}$ is simply $\beta \boldsymbol{n}$ and $\|\beta \boldsymbol{e}\| \leq \beta\epsilon$, hence discrepancy becomes

$$\|\alpha A\tilde{\boldsymbol{x}} - \beta \boldsymbol{b}\| = \beta\eta\epsilon,$$

which is satisfied when

$$\tilde{\boldsymbol{x}} = \beta/\alpha \boldsymbol{x}, \quad \text{and} \quad \tilde{\mu}^i = \alpha^2/\lambda^2 \mu^i.$$

Parameter selection

Continued

Choosing ω^i to "undo" scaling implies $\tilde{\mu}^i = \mu\tilde{\omega}^i$.

Need $\tilde{\omega}^i$ s.t. $\tilde{\omega}^i = \alpha^2/(\lambda^i)^2\omega^i$ but

$$\omega^i = \frac{1}{\|D\mathbf{x}^i(\mu^i)\|} \sim \frac{\alpha^3}{(\lambda^i)^2\beta}$$

Easy fix

$$\omega^i = \frac{\|\mathbf{x}^i(\mu^i)\|}{\|D\mathbf{x}^i(\mu^i)\|} \sim \frac{\beta}{\alpha}\frac{\alpha^3}{(\lambda^i)^2\beta} = \frac{\alpha^2}{(\lambda^i)^2}$$

Alternative

$$\omega^i = \frac{\|A\mathbf{x}^i(\mu^i)\|}{\|DA\mathbf{x}^i(\mu^i)\|}$$

Results

Results

Setup

1D

- Problems from Regularization tools (Hansen, 1994)
- $\epsilon = 0.01\|\boldsymbol{b}\|$, $\eta = 1.01$.
- Differential regularization operator and orthogonal projection.

2D

- 412x412 image `blurred` with $\sigma = 5$ and half-bandwidth 11.
- Total-variation type regularization based on the Perona–Malik diffusion equation.
- $\epsilon = \mathbb{E}[\|\boldsymbol{n}\|] = 0.05\|\boldsymbol{b}\|$, η s.t. $\|\boldsymbol{n}\| \leq \eta\epsilon$ in 99.9% of the cases.

Numerical experiments

Results

Table: Median error of 100 runs for different problems.

Problem	Single	Multi	Ratio
Baart	$1.73 \cdot 10^{-1}$	$3.04 \cdot 10^{-2}$	$1.76 \cdot 10^{-1}$
Deriv2-1	$2.25 \cdot 10^{-1}$	$3.81 \cdot 10^{-3}$	$1.69 \cdot 10^{-2}$
Deriv2-2	$2.29 \cdot 10^{-1}$	$1.98 \cdot 10^{-2}$	$8.65 \cdot 10^{-2}$
Deriv2-3	$4.36 \cdot 10^{-2}$	$4.32 \cdot 10^{-2}$	$9.91 \cdot 10^{-1}$
Foxgood	$3.28 \cdot 10^{-2}$	$2.42 \cdot 10^{-3}$	$7.38 \cdot 10^{-2}$
Gravity-1	$3.68 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	$4.89 \cdot 10^{-1}$
Gravity-2	$5.54 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$	$7.22 \cdot 10^{-1}$
Gravity-3	$1.01 \cdot 10^{-1}$	$9.13 \cdot 10^{-2}$	$9.06 \cdot 10^{-1}$
Heat-5	$1.01 \cdot 10^{-2}$	$1.01 \cdot 10^{-2}$	$1.00 \cdot 10^{+0}$
Heat-1	$8.51 \cdot 10^{-2}$	$8.79 \cdot 10^{-2}$	$1.03 \cdot 10^{+0}$
Phillips	$2.36 \cdot 10^{-2}$	$2.07 \cdot 10^{-2}$	$8.75 \cdot 10^{-1}$
Shaw	$1.12 \cdot 10^{-1}$	$1.12 \cdot 10^{-1}$	$1.00 \cdot 10^{+0}$

Numerical experiments

Results

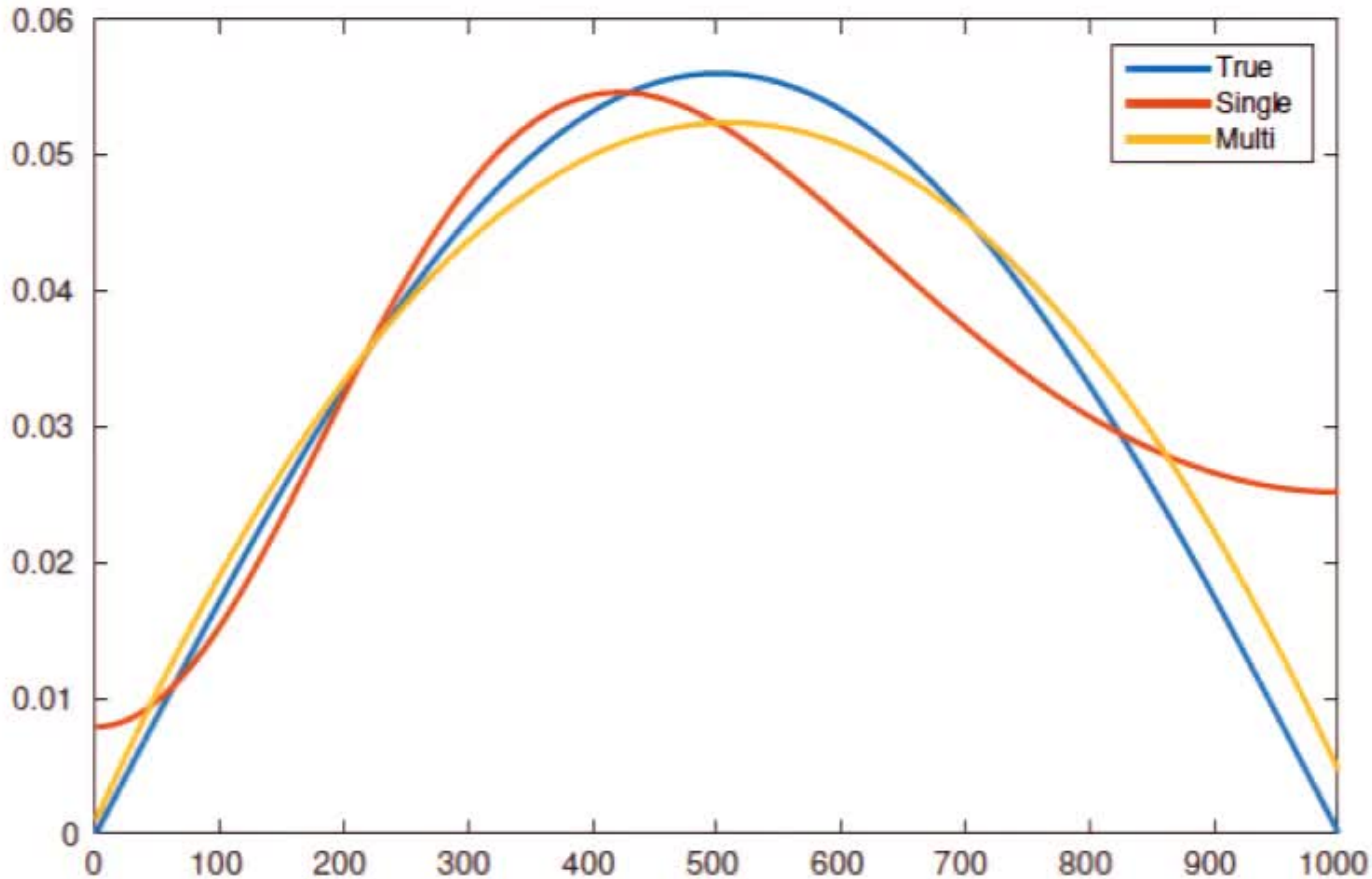

Figure: Approximating the solution from baart (blue) with single (red) and multidirectional (yellow) subspace expansion.

Numerical experiments

Gaussian blur

Figure: Blur test case; original (left), blurred and noisy (middle), and reconstructed (right).

Numerical experiments

Gaussian blur convergence

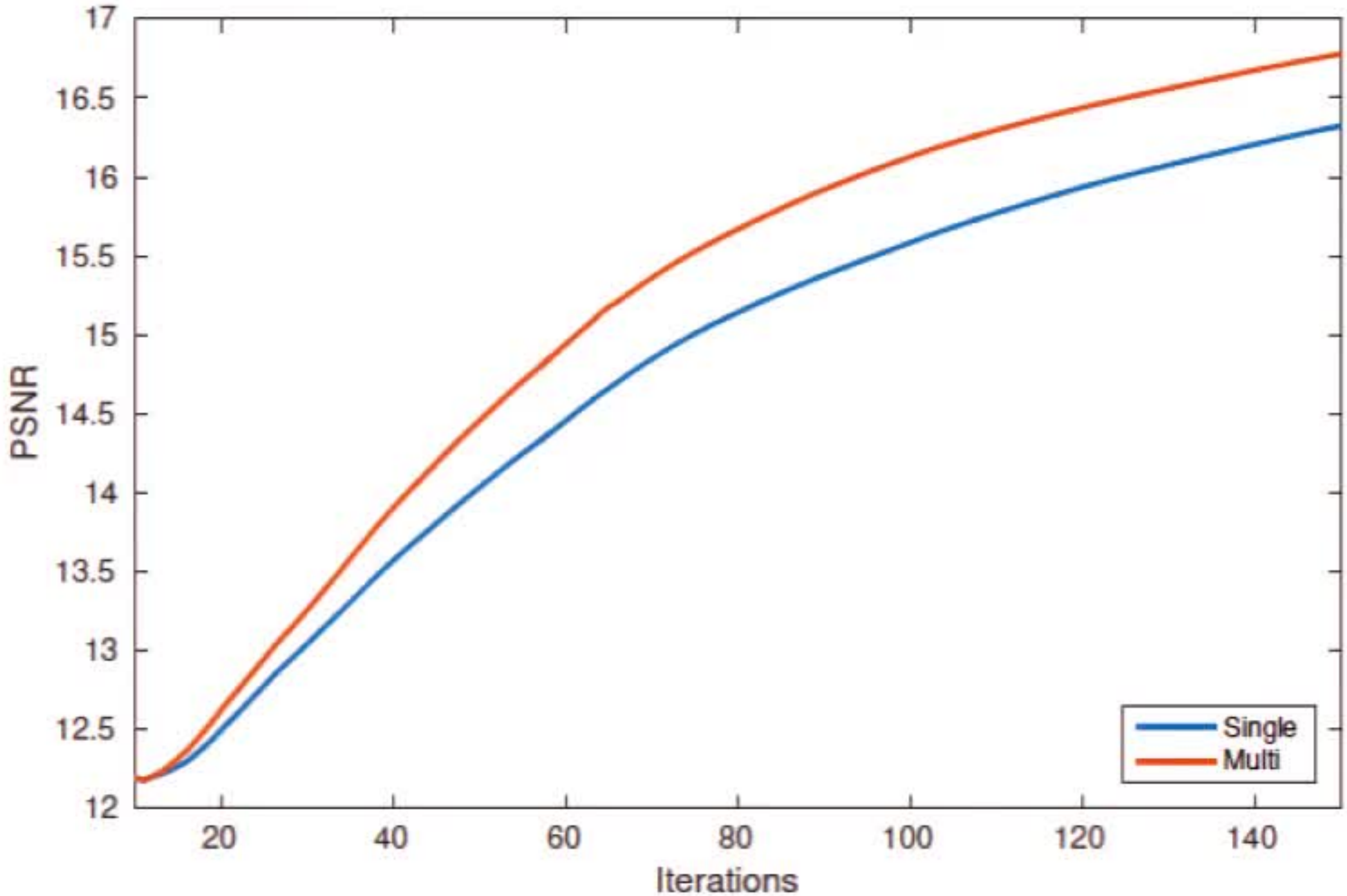

Figure: Blur test case; PSNR vs iteration number and single (blue) vs multidirectional (red) subspace expansion.

Numerical experiments

Gaussian blur performance

Table: The number of matrix-vector products and wall clock time used by the different methods.

Method	Total	A	A*	L	L*	Time (s)
Single	599	150	150	150	149	46.5
Multi	857	279	150	279	149	56.9
Parity	629	203	112	203	111	37.3

Q: Why can we beat "Single" even though we use more MVs?
A: Because we can exploit blocking operations.